Data Driven Multispectral Image Registration Framework

A Thesis Submitted to the
College of Graduate and Postdoctoral Studies
in Partial Fulfillment of the Requirements
for the degree of Master of Science
in the Department of Computer Science
University of Saskatchewan
Saskatoon

By
Rahat Yasir

Permission to Use

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

Abstract

Multispectral imaging is widely used in remote sensing applications from UAVs and ground-based platforms. Multispectral cameras often use a physically different camera for each wavelength causing misalignment in the images for different imaging bands. This misalignment must be corrected prior to concurrent multi-band image analysis. The traditional approach for multispectral image registration process is to select a target channel and register all other image channels to the target. There is no objective evidence-based method to select a target channel. The possibility of registration to some intermediate channel before registering to the target is not usually considered, but could be beneficial if there is no target channel for which direct registration performs well for every other channel.

In this paper, we propose an automatic data-driven multispectral image registration framework that determines a target channel, and possible intermediate registration steps based on the assumptions that 1) some reasonable minimum number of control-points correspondences between two channels is needed to ensure a low-error registration; 2) a greater number of such correspondences generally results in higher registration performance.

Our prototype is tested on five multispectral datasets captured with UAV-mounted multispectral cameras. The output of the prototype is a registration scheme in the form of a directed acyclic graph (actually a tree) that represents the target channel and the process to register other image channels. The resulting registration schemes had more control point correspondences on average than the traditional register-all-to-one-target-channel approach. Data-driven registration scheme consistently showed low back-projection error across all the image channel pairs in most of the experiments. Our data-driven framework has generated registration schemes with the best control point extraction algorithm for each image channel pair and registering images in a data-driven approach. The data-driven image registration framework is dataset independent, and it performs on datasets with any number of image channels. With the growing need of remote sensing and the lack of a proper evidence-based method to register multispectral image channels, a data-driven registration framework is an essential tool in the field of image registration and multispectral imaging.

Acknowledgements

I would like to thank my supervisor, Dr. Mark G. Eramian, for his guidance, support, and encouragement. I am thankful to Canadian First Research Excellence Fund for funding this research project. I am thankful for all the support from Plant Phenotyping and Imaging Research Center (P2IRC) of University of Saskatchewan. Special thanks to Dr. Ian Stavness and Dr. Kevin Stanley for their support.

To everyone at IMG - Imaging Multimedia and Graphics Lab of University of Saskatchewan.

Contents

List of Tables

List of Figures

List of Abbreviations

LOF	List of Figures
LOT	List of Tables
SIFT	Scale-invariant feature transform
SURF	Speeded up robust features
BRISK	BRISK: Binary Robust Invariant Scalable Keypoints
ORB	Oriented FAST and Rotated BRIEF
CP	Control Point
BPError	Back Projection Error
RT	Run Time
KNN	k-nearest neighbors
MST	Maximum Spanning Tree
BFMatcher	Brute Force Matcher
Nir	Near Infrared Channel
Rdg	Red Edge Channel

Chapter 1

Introduction

1.1 Problem Description

Many multispectral cameras use different physical cameras with different physical sensors for sensing different channels which causes spatial misalignment of the image channels due to their physical displacement (see Figure 1.1 left image). The traditional approach for multispectral image registration is to designate one channel as the target channel and register all the other image channels to the target. Currently, there is no objective evidence-based method to select a target channel. Registration of two image channels is achieved by inferring the necessary transformations from a set of control-point correspondences that pair identical points in the scene in each of the two images. Generally, the more such correspondences can be found, the better the alignment of the channels will be. The lack of an objective evidence-based method to select the target channel that is providing maximum average control point correspondences across all image channel pairs affects multispectral image registration performance. Moreover, the possibility of registration to some intermediate channel to the target is not usually considered but could be beneficial if there is no target channel for which direct registration performs well for every other channel.

1.2 Motivation

The previously stated problem in problem description is important because multispectral imaging is widely used in the field of remote sensing for surveillance [8], oceanography [26], geology [26], glaciology based studies [26], military intelligence [8] and precision agriculture based applications [17]. It is popular for analysis of food crops because various visible, red-edge, and near-infrared bands are important for assessing crop health and vigor. Recently crops are more frequently imaged using small multispectral camera devices mounted on unmanned aerial vehicles (UAVs). In particular, plant breeders want to use multispectral cameras to extract phenotypic traits from aerial images of field trials captured using unarmed aerial vehicles (UAV) or drones equipped with multispectral cameras. Farmers are using near-infrared sensors with their drones to generate NDVI image indexes to get pheno-typic traits of different crop lands [17]. The demand for multispectral cameras is rising with the increasing variety usage of remote sensing applications.

1.3 Objectives

The objective of this research is to design a data-driven approach for multispectral image registration which will solve the problem of lack of an objective evidence-based method to select the target channel that is providing maximum average control point correspondences across all image channel pairs affects multispectral image registration performance and also will consider the possibility of registration to some intermediate channel to the target if there is no target channel for which direct registration performs well for every other channel.

The data-driven framework will use a subset of the entire dataset to be registered to determine the average number control points that can be found between each pair image of channels for each control point algorithm in some candidate set of control point extraction algorithms. A registration scheme is then determined that describes the target channel and the process to register all the other image channels will be registered to the target channel using the best control point extraction algorithm for each pair of channels. The framework may allow intermediate registration steps for image channels if direct-to-target registration produces fewer control point correspondences on average than intermediate steps.

Our framework is tested on five image datasets captured with UAV-mounted multispectral cameras. The proposed framework is dataset-independent and is compatible with images having any number of channels. We evaluated its effectiveness relative to the traditional approach on five datasets: 1) Canola Crop Breeding Data, 2) Greenhouse Canola Plants Data, 3) UAV Mixed-Crop Dataset-1, 4) UAV Mixed-Crop Dataset-2, and 5) TokyoTech dataset. Figure 1.1 demonstrated the misaligned and registered RGB images of canola crop dataset before (left) and after (right) registration.

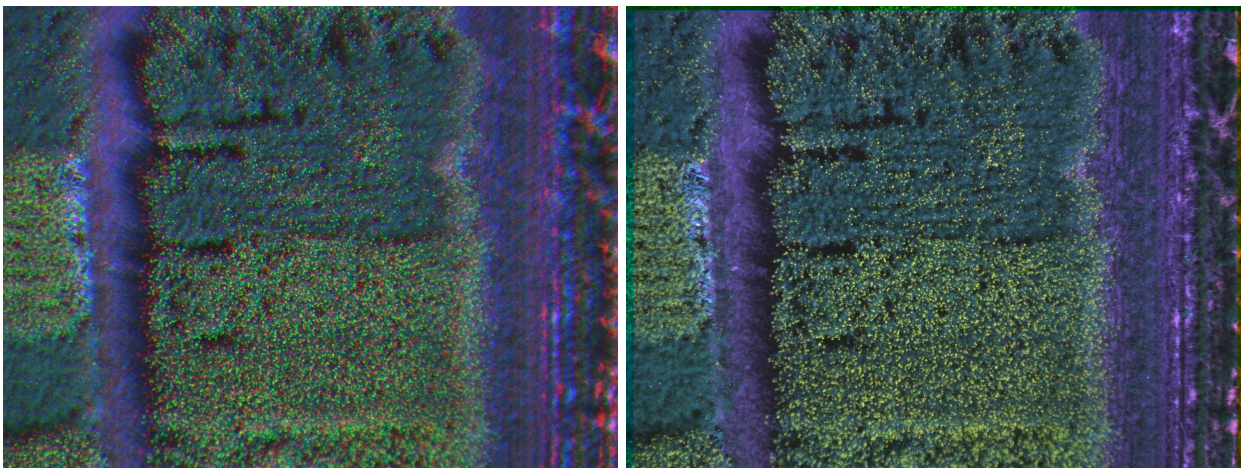

Figure 1.1: Misaligned RGB image of Canola crop dataset(left) and registered RGB image of canola crop dataset using data-driven multispectral image registration framework (right)

1.4 Outline

This thesis is organized as follows, Chapter 2 presents image registration preliminaries, basic definitions, and techniques used in the thesis. Chapter 3 presents a structured literature review of existing image registration works, image registration performance evaluation methodologies and existing work on multispectral image registration, limitations, and scope of improvements. Also, existing image registration standard control point extraction algorithms are reviewed in chapter 3. Chapter 4 presents the new data-driven multispectral image registration framework along with an overview of data collection and datasets. Chapter 5 presents the evaluation of the framework and compares the results to standard direct-to-target registration. We have discussed the experimental setup process, control point extraction algorithms performance evaluation, dataset wise registration scheme performance, performance statistics of different approaches and failure cases. Chapter 6 provides a comprehensive discussion of the overall research and findings. Chapter 7 a summary of research contributions and a discussion of future work.

Chapter 2

Image Registration Preliminaries

2.1 Image registration fundamentals

Image registration is the process of aligning two or more images of the same scene. This process involves designating one image as the target image, and applying geometric transformations or local displacements to the other images so that they align with the target. Images taken using the same camera and same scene but from different view point are misaligned. Misalignment can also result from lens and sensor distortions or differences between capture devices.

In this section of the thesis we will review fundamental concepts related to image registration framework. We will use these topics throughout the thesis.

2.1.1 Image registration process

The image registration process involves designating one image as a target image and applying geometric transformations to the other images so that they align with the target image. Hong and Zhang described image registration as a problem of geometrically aligning images taken from different viewpoints. They have mentioned that image registration consists of four steps [14],

I. Feature extraction (control-points) between two images (target and sensed image)

II. Feature matching which establishes control-point correspondences between two images

III. Mapping function construction

IV. Image transformation

I. Control-point Extraction

Image registration is geometrically transforming and resampling the sensed image according to the target image. Image registration control point extraction algorithms find control-points by analyzing the magnitude and direction of intensity changes in local image neighbourhoods to detect high-contrast corners, edges and intensity patterns of the image [23]. Each of the control point is consists of a feature vector and each feature vector is a binary string that encodes information about the magnitude, pattern, dominant direction of gradients of each control point. Figure 2.1 image is showing the detected independent control-points of an

image. Control points are surrounded by a circular window centered on them. The radius of each circular window represents the scale of each control point. To find control-points, images are downsampled in multiple scales, and we check for potential control-points at each scale. All the control-points of different scales are combined as the final output. The larger the radius of the circle, the larger the scale was that the control point was detected at. The line from the center to the edge of the circle is considered as the orientation of the control point. The lines of control point orientation represents the most dominant orientation of the gradient angles of that circular window.

Figure 2.1: Independent control-points in an image.

2.1.2 II. Control-point Matching

A brute force matcher matches each value of one set with all the values of another set and returns the closest match. The principle of brute force matcher in image registration is to take one set of control-points and match it with the control-points of another set one by one according to their Euclidean distance. As each control point has a binary string that encodes image intensity pattern information considered as feature

vector. The binary encoding allows very fast comparison of control-points from different images using brute force matching. A brute force matcher matches control-point correspondences of two images. The brute force matcher works as a string matcher between two sets of binary strings. Figure 2.2 shows two sets of control-points with their centre location i and j. We are finding the closest match of the first control point location from the second set of control-points. The brute-force matcher returns the location of the closest match (i, j) for each control point in the first set. The drawback of Brute-Force matcher is, it is computationally expensive due to checking a large number of control-points and the complexity of brute force matching is O(n·m) where n is the number of control-points and m is the size of the feature vector for each control point. Control point extractor algorithms may find thousands of independent control-points from an aerial image. D. Lowe's [24] ratio test is applied on control-point correspondences to eliminate false matches. Figure 2.3 image shows control-points matched between two different images.

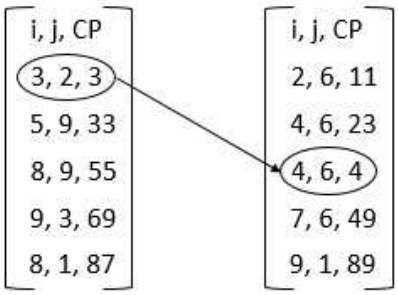

Figure 2.2: Brute Force matcher based matching between two sets of points. (3,2,3) in set one best match with (4,6,4)

2.1.3 III. Mapping Function Construction

Homography matrix generation is the mapping function construction of image registration from control-point correspondences. A homography is a transformation (a 3×3 matrix) that maps the control-points in one image to the corresponding control-points in another image [11]. Image transformation through mapping function happens if both the images are of the same scene but from different angles, from the same camera but from different viewpoints and depends on the scene structure. The minimum number of control-point correspondences required to obtain a homography matrix is four to have views of four planes from two different view points. It is always better to have more than four control-point matches. We can represent a homography matrix as,

$$\mathrm{H}=\begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix}$$

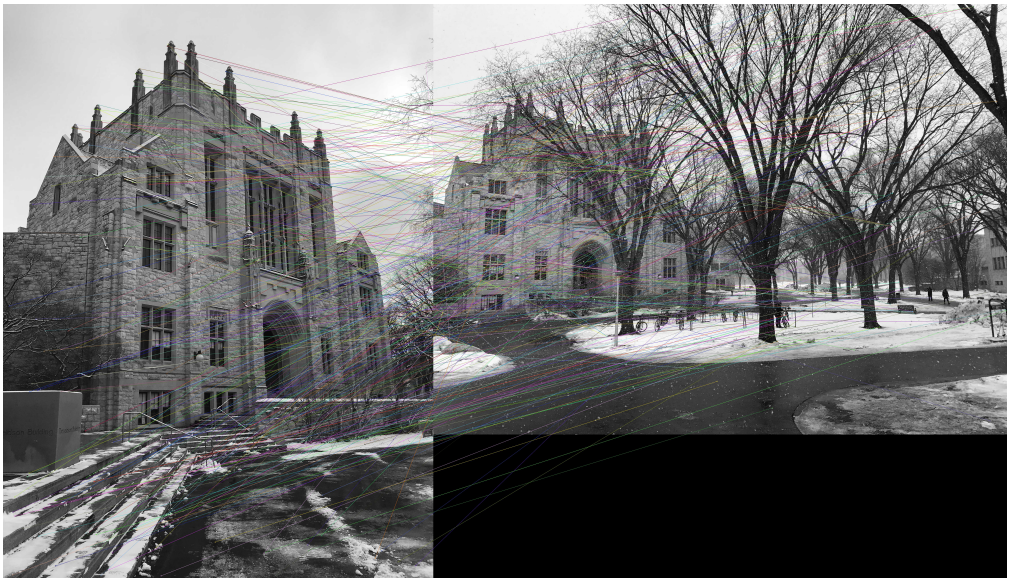

Figure 2.3: Control-points match between two images (right image)

Suppose, first set of control-points of target image is $[x_1, y_2]$ and second set of control-points representing any other image is $d[x_1, y_2]$. The mapping function for homography constructs in the following way,

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = H \begin{bmatrix} x_2 \\ y_2 \\ 1 \end{bmatrix} = \begin{bmatrix} h_{00} & h_{01} & h_{02} \\ h_{10} & h_{11} & h_{12} \\ h_{20} & h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix}$$

2.1.4 IV. Image Transformation

Final step of image registration process is image transformation according to the mapping function of step III. For transforming images according to the target image, target image (a) (figure 2.4) is warped and use the result in other image (b) using the mapping function H calculated in step III using the target image (a) and other image (b) to get transformed image (c). The process of image transformation happening here is a warp perspective where the perspective transformation is appliedening accordi the warped target image. The pPerspective trnsformation is transformsing frone project space to another from two different centers of projection and image warping mapsis mapping points of onth another without changing anything. The pPerspective transformationworks for a set of control-points where the homography matrix H is applied on control-points of the sensed image (b) to get the transformed image (c).

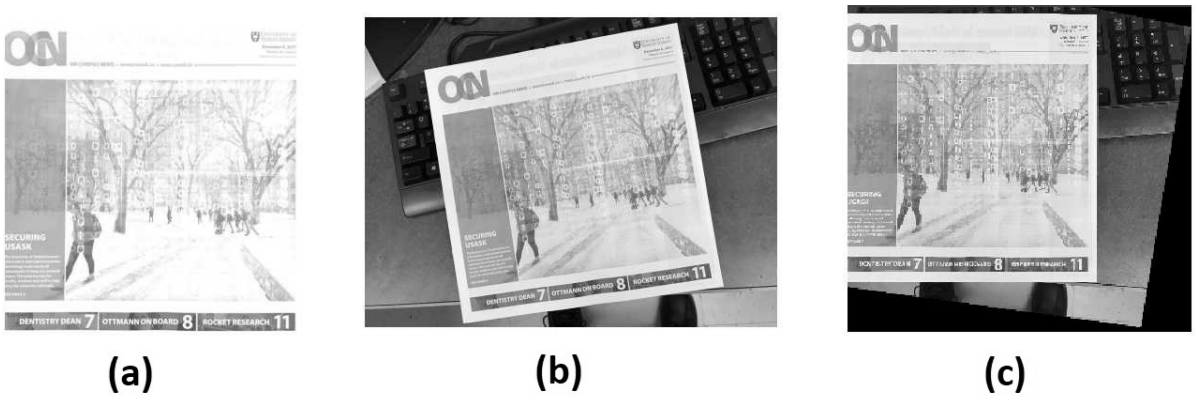

Figure 2.4: Sample image transformation example using homography matrix. Image (b) is transformed to (c) using target channel (a) (Photo content: On campus news, USask, January, 2018)

2.1.5 Performance Matrices For Registration

Back projection error is used as the performance metric for image registration. It represents how well the control-point locations of an image align with the control-point locations of a new image [1]. Back projection error is one of the key ways to measure the success of image transformation. Figure 2.5, we have two different images. We want to transform image (b) using the target image (a). Here, we are using two control point correspondences in the image pair out of many control point correspondences. We have transformed image (b) according to the target image (a) and got registered image (c). We are taking the pixels of target image control-points and using them in our registered image to find the distance between registered image control-points and target image control-points. In our diagram, blue boxes are the target image control-points, and orange boxes are registered image control-points, and the centre distance between these two sets of control-points are considered as the back-projection error. Smaller back-projection error indicates better image registration performance. The equation to calculate back projection error is,

BP (I,J)= $\sum_{x_i, x_j} d^2(X_I, HX_j)$

where H is the homography matrix and $[X_I, X_J]$ is overall control-point correspondences between image I and J. Image I is the target image.

2.2 Graph algorithms in data-driven framework

Two graph algorithms that we have used in our data-driven registration framework are described in this section.

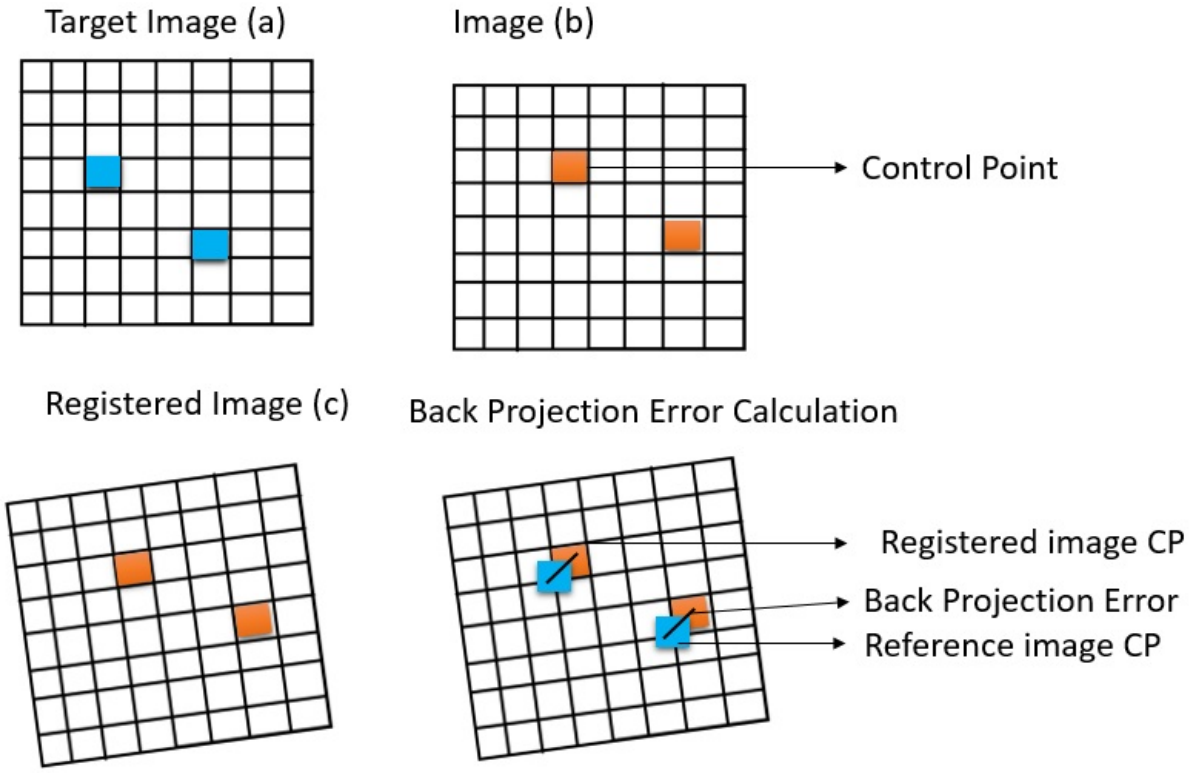

Figure 2.5: Back projection error calculation demonstration.

2.2.1 Maximum spanning tree

A spanning tree T of a weighted undirected graph G is a sub-graph that is a tree which includes all of the vertices of G, with minimum / maximum possible number of edges. A graph can have several spanning trees but a graph that is not connected does not have a spanning tree. A maximum spanning tree (MST) is a subset of the edges of a connected, edge-weighted undirected graph that connects all the vertices, without any cycles and with the maximum possible sum of edge weights. Maximum spanning tree is a greedy algorithm. The concept of maximum spanning tree is similar to the minimum spanning tree where the total weight of all the edges is maximum instead of the minimum.

Maximum spanning tree construction has three steps,

1. Sort the graph edges in ascending order of their weights.
2. Add edges to growing maximum spanning tree from the edge with maximum weight to the minimum weight.
3. Add edges which do not form a cycle and connect all the disconnected components.

We have considered Kruskal's algorithm to construct maximum spanning tree in this research. Kruskal's algorithm treats the graph as a forest and every node has as an individual tree. A tree connects to another

only if it has the maximum cost among all the available options and does not violate maximum spanning tree properties [25]. Figure 2.6 - left graph represents a simple graph of 5 different nodes and right graph is the maximum spanning tree of the left graph. We have only considered the largest edge values in the graph to construct the MST, there is no connected cycle in it and all the 5 nodes are connected here in the spanning tree. The sum of the edge weights of the maximum spanning tree is 1341 which is the maximum for any spanning tree of that graph.

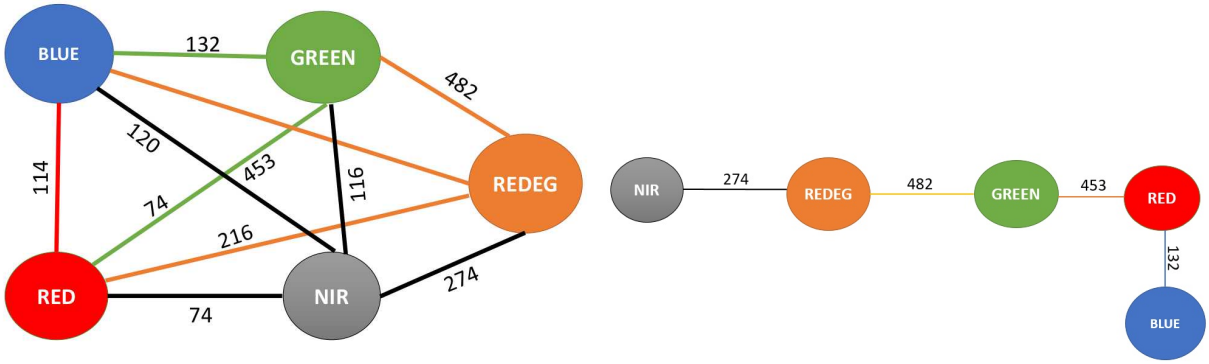

Figure 2.6: A graph (left) and it's maximum spanning tree (right).

2.2.2 Floyd-Warshall Algorithm

The Floyd-Warshall algorithm is an algorithm for finding shortest paths in a weighted graph with positive or negative edge weights but with no negative cycles [13]. Floyd-Warshall calculates the shortest distance between each pair of nodes in a graph where distance $d(u, v)$ between two vertices u and v of a finite graph is the minimum length of the paths connecting them. If the graph on the right of figure 2.6 is the input and the edge weight is considered to be 1, the output is table 2.7.

Channels	Near-Infrared	Red-Edge	Green	Red	Blue	Summed Weights
Near-Infrared	∞	1	2	3	4	10
Red-Edge	1	∞	1	2	3	7
Green	2	1	∞	1	2	6
Red	3	2	1	∞	1	7
Blue	4	3	2	1	∞	10

Figure 2.7: Tree root calculation using the Floyd-Warshall algorithm.

Chapter 3

Literature Review

In this thesis, we have followed structured literature review methodology to conduct a comprehensive literature review. Our structured literature review process is divided into three phases, objectives, keywords for comprehensive search, and inclusion criteria. The objectives of this review are to answer these questions:

1. What are the standard control point extraction algorithms?
2. What methodologies have been used to register multispectral images?
3. What are the ideal performance evaluation process for image registration ?

The second phase of the structure literature review process is to select search keywords to be used to locate relevant publications. Keywords that we have used to search related kinds of literature are,

- Multispectral image registration
- Image registration in remote sensing
- Image aligning problem
- Optimal image registration
- Drone image registration
- Image registration validation
- Data-driven image registration
- Control point extraction algorithms

We have used these keywords to find out publications related to all of our research questions.

The third phase of the structure literature review process is to collect publications using the keywords for comprehensive search to get answers of our objectives. The inclusion criteria that we have followed to collect publications are,

- Conference/journal publications post-1995
- Peer-reviewed journals and conferences

- Proposes a new algorithm rather than just using it in a system/pipeline
- Reviews evaluation methods for image registration algorithms more generally
- Published in one of the key publications in this area - Computer Vision and Image Understanding
- IEEE Transaction on Image Processing, Pattern Recognition, Medical Imaging, Pattern Analysis, Pattern Recognition Journal, Computers and Graphics, International Journal of Computer Vision, Image and Vision Computing, Image Features and Descriptors, Graphical Models and Image Processing, ICCV - International Conference on Computer Vision, ECCV - European Conference on Computer Vision, Computer Vision Pattern Recognition, International Conference on Image Processing, British Machine Vision Conference, Asian Conference on Computer Vision, Canadian Conference on Computer and Robot Vision, Remote Sensing Journals etc.

We have also considered a few publications related to our research that does not fall under the inclusion criteria as well where those literatures are closely related to our work.

We have divided the multispectral image registration literature review into three phases. First we will review standard control point extraction algorithms to select a set of candidate control point extraction algorithms for our data-driven image registration framework. Then, we will review existing multispectral image channel registration approaches. Finally, we will review standard evaluation process to evaluate image registration performance.

3.1 Control Points for Image Registration

A significant amount of work has been done in the field of image registration. Most of the works on control point algorithms in the last 15 years are focused on increasing the number of control-points that can be extracted in different situations and datasets. Control point extraction algorithms that are widely used in the field of image registration are SIFT, SURF, BRISK, and ORB.

3.1.1 SIFT

Scale-invariant feature transform (SIFT) is a control point extraction algorithm that detects and describes local features in an image. The main objective of SIFT is to identify control-points where there exist characteristics that are invariant to scaling and rotation [32]. SIFT consists of four stages [15]:
(1) Scale-space extrema detection
(2) Control point localization
(3) Orientation assignment
(4) Control point extraction

At first, SIFT scans the image to locate potential control-points. Control points are invariant to image scale and rotation. Based on the difference of Gaussian scales, insignificant control-points are rejected and edge response is eliminated. To obtain potential control-points that invariant to rotations, an orientation histogram is formed using difference of gaussian function around the control point [15]. SIFT constructs a feature vector by considering the direction of control-points and the gradient strength of control-points are maximal. According to D. Lowe, features generated by SIFT are highly distinctive, in the sense that a single feature can be correctly matched with a large number of images from many images [24].

3.1.2 SURF

The SIFT extraction algorithm partly inspired the Speeded-Up Robust Features (SURF) extraction algorithm. SURF uses an integer approximation of the determinant of Hessian blob detector to find control-points from an image. The determinant of Hessian blob detector can be computed with three integer operations using a precomputed integral image [7]. This control point extraction algorithm is based on the sum of Haar wavelet components around the control point. SURF focuses on scale and in-plane rotation-invariant detectors and descriptors. To ensure control-points of the image are scale invariant, SURF converts the image into coordinates using a multi-resolution pyramid technique and obtains an image with reduced bandwidth. This achieves a special blurring effect on the image and creates scale-space. Scale-space makes control-points scale invariant [7].

3.1.3 BRISK

Binary Robust Invariant Scalable Key Points (BRISK) is another image control point extraction algorithm that was designed to improve upon well-established algorithms such as SIFT and SURF. Similar to SIFT, the BRISK algorithm consists of four major stages,
(1) Scale-space control point detection
(2) Control point filtering and sub-pixel localization
(3) Orientation assignment
(4) Descriptor generation
To find out potential control-points, the first step is to downsample the original image to build a pyramid scale-space [28]. Control points are selected using the FAST 9-16 detector from the image pyramid. Control points are filtered by a 3D non-maxima suppression within the scale-space pyramid. To match the orientation of the control-points of image pairs, gradient distance is calculated using Daisy descriptor [28]. This binary control point extraction algorithm does not describe the absolute difference between two control-points. It returns the monotonic grey value changes between the scenes.

3.1.4 BRIEF

BRIEF stands for Binary Robust Independent Elementary Features is a feature descriptor. BRIEF represents a faster method for feature calculation, matching and a high recognition rate, unless there is a large in-plane rotation. SIFT uses 128 dimensional vector and SURF uses 64 dimensional vector for control point features. All these dimensions are not necessary for matching and it requires larger memory and processing time for matching of those vectors. BRIEF takes a smoothened image patch and selects a set of location pairs to do pixel intensity comparisons on those location pairs using Hamming distance. Location pair distances are represented as 1 or, 0 to construct n-dimensional bitstring [10].

3.1.5 ORB

Oriented FAST and Rotated BRIEF (ORB) is a very fast binary control point extraction algorithm based on BRIEF. It is a fast robust local feature detector. It is based on the FAST control point extraction and the visual control point correspondence of BRIEF (Binary Robust Independent Elementary Features) [27]. It aims to provide a fast and efficient alternative to SIFT. To get potential control-points, ORB constructs a scale pyramid of the original image by downsampling it. Control points are detected from each scale of the pyramid using FAST detector [15]. The Hassic corner measure sort all the potential control-points and top control-points are chosen based on a threshold value. To calculate rotation invariant of control point correspondences from image pairs, first-order moments are used to compute the local orientation of each control point through intensity centroid magnitudes. ORB control point extraction algorithm describes the features of the input image in a binary string instead of a vector.

3.1.6 PCA-SIFT

PCA stands for principal component analysis which is a dimensionality reduction algorithm. Luo et. al. has modified SIFT and introduced PCA instead of histogram to normalize gradient patches [16]. By using a dimensionality reduction algorithm, the feature vector generated from PCA-SIFT is significantly smaller than the standard SIFT feature vector. PCA-SIFT uses Euclidean distance to determine whether the two vectors correspond to the same keypoint in different images. According to [16], PCA-SIFT has fewer components to process and this requires less storage which results faster matching and significant space benefits.

3.1.7 FREAK

The Fast Retina Keypoints (FREAK) control point extraction system was inspired by the human visual system. FREAK uses circular sampling grids to compare pairs of pixel intensities where there is higher density of sample points near the center and exponentially decreasing density further away. Each of the circle represents the standard deviations of the Gaussian kernels applied to the corresponding sampling points.

Different Gaussian kernels with respect to the log-polar retinal pattern leads to a better performance than general circular grid templates. Methodologies to find out control-points and matching are similar to BRISK and ORB, the difference being the exponential change in size of the circular grid and the overlapping receptive fields. Each control point in FREAK is smoothed to reduce the effects of noise. Human eyes look around with unique discrete movements named cascades. Almost 90% of the incorrect matched points are discarded with FREAK's first 16 bits and these 16 bits are chosen for the first cascade. The last cascade usually selects the location of objects interest despite light and perspective changes observed in an image [6].

3.1.8 Discussion

Hong and Zhang observed that a large number of control-points are required to register high resolution images and that more control-points increase the processing time [14]. Many researchers have concluded that the SIFT control-points have excellent performance [15]. Leuteneggar et al. showed that BRISK control-points exhibit adaptive, high-quality performance with low computational cost [22]. Bay et. al. claimed that the performance of SURF control-points is comparable to or better than previously proposed schemes with respect to repeatability, distinctiveness, and robustness, with comparatively low computational cost [7]. ORB control-points are rotation invariant and resistant to noise [27]. Luo et al. compared the performance of SIFT, SURF, and PCA-SIFT. From their experiments, SIFT showed stable and consistent performance but, it was the slowest control point extraction algorithm among all the candidates. SURF proved to be faster than SIFT and PCA-SIFT with as good performance as SIFT [16]. In Işik and Özkan's comparative study of different control point algorithms, SIFT showed best performance in terms of the number of control-points correspondences [15]. In the same study, SURF had the best performance in terms of runtime.

The literature shows that SIFT, SURF, BRISK, and ORB are standard control point algorithms in the field of image registration. We will use these control point algorithms to design our data-driven image registration framework.

3.2 Multispectral image registration

Multi-spectral imaging is becoming popular with the increasing use of remote sensing applications. Due to hardware limitation of some widely used multispectral cameras like MicaSense (Micasense Inc. Seattle, WA, USA) and Sequioa SenseFly (SenseFly SA, Switzerland), multispectral image registration is a very important issue to work on in the field of computer vision. Very few researchers have worked on an infrared and visual channel based multispectral image registration problem however, no multispectral image registration method has been established to address the limitations of multispectral cameras.

Firmenich et al. showed that it is difficult to find a sufficiently large number of control point correspondences between visible (e,g, red, green, blue) and near-infrared image channels [12]. They proposed gradient

direction invariant SIFT (GDISIFT) where material reflectance can reverse the direction of gradients by 180 degrees between the visible and NIR channels. Their results are data and image content-dependent. They have demonstrated their solution on very few images and their solution only worked when there was a large intensity difference between objects.

Zhiguo and Yang proposed modified SIFT control-points to avoid incorrect correspondences of SIFT control-points between images [31]. For registration between near-infrared, they observed high numbers of incorrect control-points correspondences. They have proposed a scale restriction criteria based gradient orientation modification process for the SIFT matcher to reduce incorrect matches when the angle of rotation is high and scale factor between 0.7 to 1. They have developed and tested their proposed method using a single set of 252 multispectral image pairs. Their proposed method is unable to increase the number of true matches and reduce the number of false matches compared to D.Lowe's ratio test [24] which was shown to eliminate 90% of false matches while discarding less than 5% of correct matches.

Jaffrey et al. proposed that multispectral imagery based registration can often be unreliable because of the divergences in scene reflectance at different wavelengths between visible and infrared imagery. They have proposed a frequency domain model for the mutual information surface around the optimal parameters to predict the behavior of the registration [19].

Teke and Temizel proposed a scale-restriction method with SURF control-points to increase the number of control point correspondences between near-infrared and visible channels and reduce the number of false correspondences [29]. Their experiment is based on 35 satellite images from the QuickBird dataset. Their solution showed a significant number of matches between the near-infrared channel and panchromatic band but, a very little effect on the near-infrared channel and red, green, blue bands.

3.3 Evaluation of registration performance

Different researchers have used different techniques to evaluate the registration process. Paula et al. emphasized the importance of proper image registration. A single pixel error may cause a complete object misalignment in the multi-temporal image sequence [21].

Wang et al. used the root mean square error (RMSE) of control-points between the registered image and the ground truth generated by an expert as the major criterion for registration error assessment [30]. They found that if root mean square error (RMSE) is more than 10 pixels, then the registration is successful.

Zitová and Flusser published a comprehensive study of image registration papers [33]. They identified three methods for evaluation of image registration accuracy:

1. Localization error - poor localization of control-points
2. Matching error - erroneous control point correspondences between images
3. Back projection error - the difference between mapping model and the actual position for the control point which is known as back-projection error (Described in Chapter 2). Cai et al. introduced affine SIFT (a

modification of SIFT) where they focused on the camera viewpoints and used the number of control-points matched as a success measure [9]. Ke and Sukthankar measured the ratio of the number of correct matches to the total number of the match to evaluate their PCA-based modified SIFT algorithm [18].

The literature places great importance on increasing the number of control-points able to be extracted as it has been shown that increased numbers of control-points generally yields more good control point correspondences and, in turn, more accurate registrations. This justifies our use of a number of control-points as the primary metric to determine the best registration scheme.

Chapter 4

Dataset and Methodology

In this chapter of the thesis, we will briefly talk about the datasets that we have collected and used in our experiments to evaluate data-driven image registration framework.

4.1 Dataset overview

We have used 5 different multispectral image datasets to design and evaluate our data-driven image registration framework. Three of the datasets are collected using different aerial drones and two of them are from fixed multispectral cameras. The dataset that we have used to design the system is the Canola crop breeding trial dataset. Figure 4.1 gives a graphical demonstration of the way we have collected Canola crop breeding multispectral image data. Datasets that we have used in this thesis are given below,

I. Canola Crop Breeding Data

II. Greenhouse Canola Plants Data

III. UAV Mixed-Crop Dataset-1

IV. UAV Mixed-Crop Dataset-2

V. TokyoTech Dataset

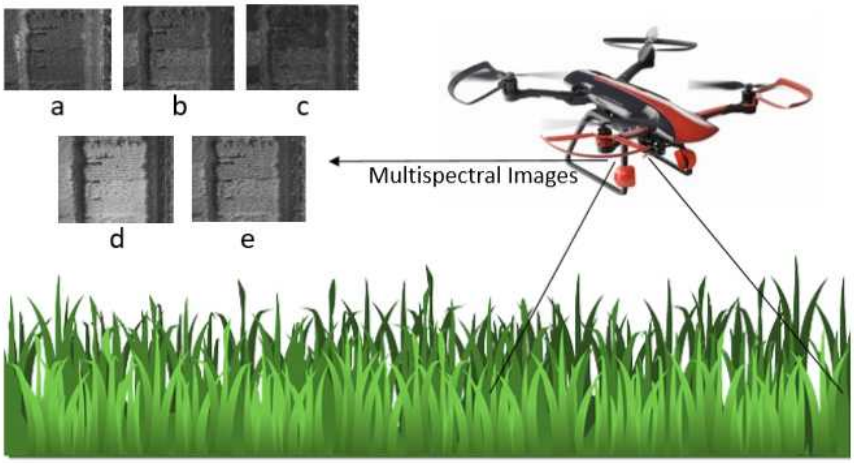

Figure 4.1: Multispectral image data collection process using drones

4.1.1 Canola crop breeding trial dataset

This dataset consists of images of a canola field for crop breeding research imaged with a MicaSense Red-Edge camera (MicaSense Inc. Seattle, WA, USA) camera mounted on a Draganfly X4P quad-copter (Draganfly Innovations Inc., Saskatoon, SK, Canada) with an actively stabilized two-axis gimble. The canola trial is located at the Agriculture and Agrifood Canadas Saskatoon Research and Development Centre, Llewellyn Road Farm (latitude 52 1053.59N longitude: 106 3010.19W) near Saskatoon, Saskatchewan, Canada. The canola field consisted of 216 individual canola plots (6 m $\times$ 2 m). Canola field breeding drone based imagery was captured during the summer 2016 growing season and the UAV was flown at an altitude of 15 m. The Red-Edge camera captures five spectral channels; red, blue, green, near-infrared and red-edge (mix of red and near-infrared). Each channel is imaged through separate optics and sensors. We have used five drone flights images and each flight has 600 images. In total, we have 3000 images in this dataset, each of 1280 $\times$ 960 pixels. Figure 4.1 demonstrates the process of collecting data using a quad-copter drone in a canola field. The cCapture rate of the MicaSense multispectral camera is 1 capture pe second of all bands in 12-bit raw image [5]. Figure 4.2 shows the sample photo of one scene (all bands) collected using MicaSense multispectral camera stacked with Draganfly drone.

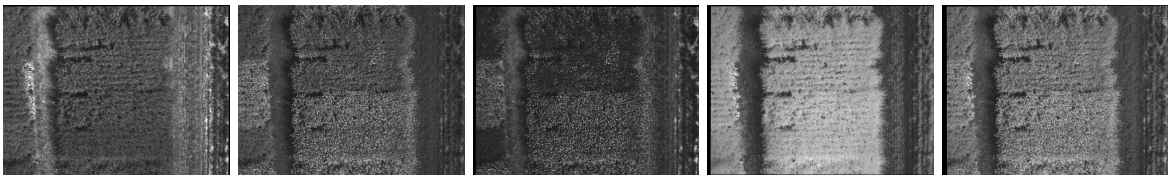

Figure 4.2: Canola crop breeding multispectral images - Blue, Green, Red, Near-Infrared and Red Edge channels from left to right.

4.1.2 Green house dataset

This dataset consists of top-down images of canola plants being grown in a greenhouse setting. A MicaSense Red-Edge camera (Mica-sense Inc. Seattle, WA, USA), as described above, was mounted on a stick in a fixed position and was used to capture still photos of greenhouse canola plants. We have used canola crop greenhouse dataset to evaluate our designed data-driven framework. This dataset is collected from the canola crop greenhouse based on Saskatoon, Saskatchewan. There are 4800 multispectral images in the dataset, each of 1280 $\times$ 960 pixels. Images were captured manually by a researcher and stored as 12-bit raw images [5].

Figure 4.3: Green house multispectral images - Blue, Green, Red, Near-Infrared and Red Edge channels from left to right.

4.1.3 UAV Mixed-Crop Dataset-1

Our third dataset is a publicly available multispectral agricultural dataset from SenseFly (SenseFly SA, Switzer-land). This dataset was used only to evaluate our completed image registration framework. UAV mixed-crop dataset-1 was collected in Switzerland and it is a publicly available multispectral image dataset [4]. The dataset images about 60 acres of cropland using Senseflys eBee Ag drone carrying a SenseFly Sequoia camera. The eBee drone is a fixed-wing drone and the camera was directly mounted to the drone which makes this a more difficult dataset to register compared with the images from the gimble-stabilized camera on the quad-copter drone in our first dataset. There are 712 multispectral images in the dataset, each of 1280×960 pixels. The images consist of four image channels- Green, Red, Near-Infrared, Red Edge. Ground resolution for the dataset is 10.62 cm/px, flight height is 112 meter. Figure 4.4 shows the sample images from UAV mixed-crop dataset-1.

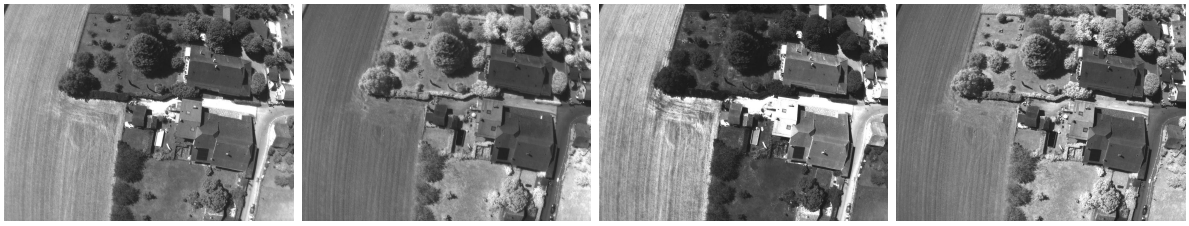

Figure 4.4: UAV mixed-crop dataset-1 images - Green, Red, Near-Infrared and Red Edge channels from left to right.

4.1.4 UAV Mixed-Crop Dataset-2

The fourth dataset that we have used to evaluate our data-driven registration framework is UAV mixed-crop dataset-2 [4]. UAV mixed-crop dataset-2 is collected in Switzerland and it is a publicly available multispectral image dataset. The dataset images about 2.57 square kilometer area of cropland using Senseflys eBee SQ drone carrying a SenseFly Sequoia camera. The eBee SQ drone is also like eBee Ag drone of the previous dataset which is a fixed-wing drone and the camera was directly mounted to the drone. There are 5260 multispectral images in the dataset, each of 1280×960 pixels. The images consist of four image channels- Green, Red, Near-Infrared, Red Edge. Ground resolution for the dataset is 12.13 cm/px, flight height is 120 meter. This dataset was used to only to evaluate our completed image registration framework. Figure

4.5 shows the sample images from UAV mixed-crop dataset-2. Key differences between UAV Mixed-Crop Dataset-1 and 2 are, flight height and ground resolution. Flight height for UAV Mixed-Crop Dataset-2 is higher than dataset 1 and edges and landmarks are less visible in Dataset-2 than Dataset-1.

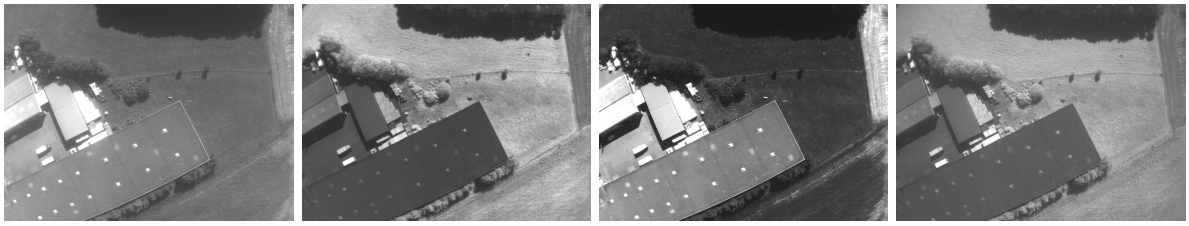

Figure 4.5: eBee SQ multispectral images - Green, Red, Near-Infrared and Red Edge channels from left to right.

4.1.5 TokyoTech Dataset

The fifth and last dataset that we have used to evaluate our data-driven registration framework is the TokyoTech multispectral image dataset [3]. Images were captured using a monochrome camera (Tokyo Institute of Technology, Tokyo, Japan) with a VariSpec liquid crystal tunable filter (VariSpec VIS). This dataset contains images of 12 scenes. There are 72 images in the dataset, each of 1824×1369 pixels. Figure 4.6 shows the sample images from TokyoTech multispectral dataset.

Figure 4.6: TokyoTech original multispectral images - Blue, Green, Red, Cy and Or channels from left to right.

The images from TokyoTech multispectral dataset are well aligned. We have rotated images randomly to misalign them. In a random order, we have rotated TokyoTech images between -3° to +3° angle. 4.7 shows images after random rotation from -3 to +3 degree.

Figure 4.7: TokyoTech misaligned multispectral images - Blue, Green, Red, Cy and Or channels from left to right.

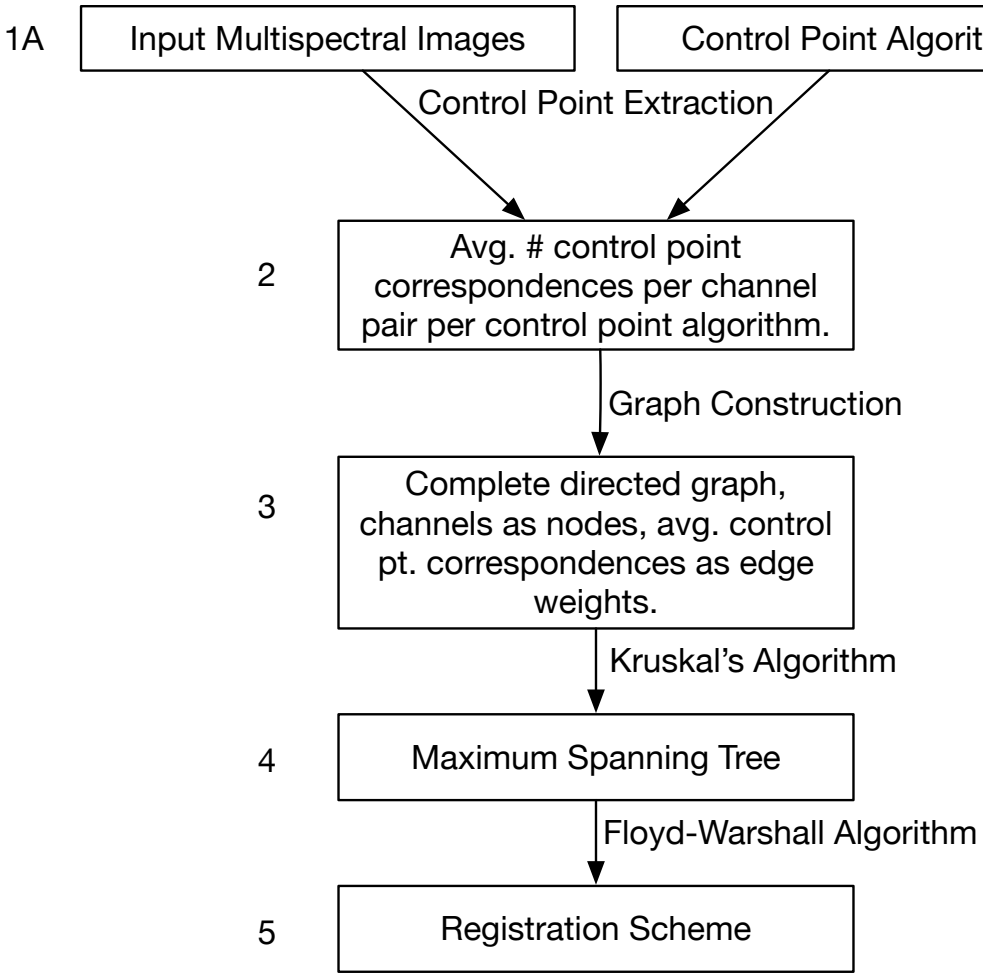

Figure 4.8: Architecture of data-driven registration framework

4.2 Data-Driven Registration Framework Architecture

The proposed data-driven image registration framework accepts as input a dataset of multispectral images (or part thereof) and a set of control point extraction algorithms. Its output is a registration scheme in the form of a directed, acyclic graph (actually a tree). The resulting registration scheme can then be applied to the entire dataset. An overview of the registration algorithm is showing in 4.12. We elaborate on each phase of the algorithm in the following sections.

4.2.1 Control Point Extraction

For each input multispectral image, we compute control point correspondences between each pair of channels for each input control point algorithm. Control point extraction algorithms that we have used in this framework are, SIFT, SURF, BRISK, ORB. Control points are extracted independently from each image

channel first. Figure 4.9 shows independent control-points obtained on a sample image using SIFT control point extraction algorithm.

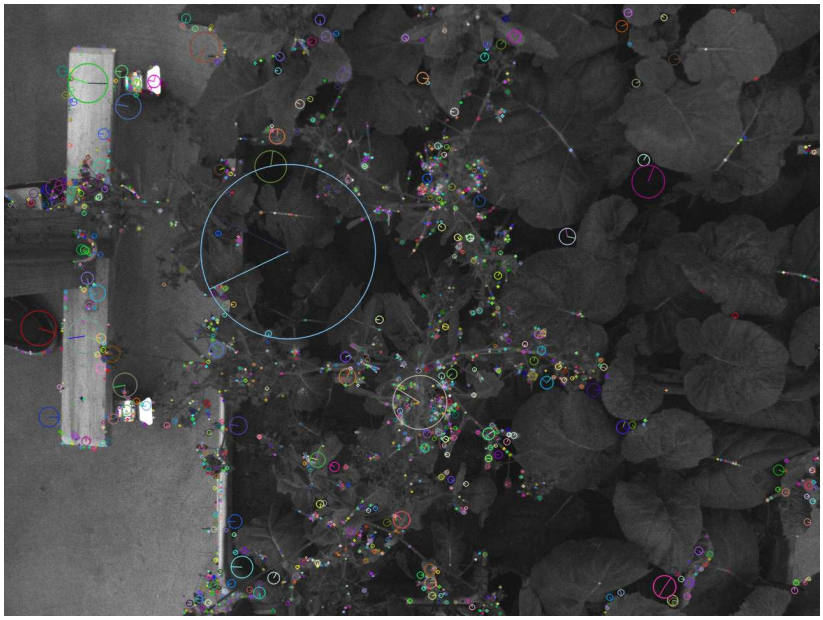

Figure 4.9: Independent control-points from SIFT control point extraction algorithm.

The average number of control-points for each pair of channels is determined for each control point algorithm (Step 2 in Fig. 4.12). Control point correspondences between a pair of images (I, J) are found for each control point extraction algorithm using brute force matching. For each control point in image I, the two most closely matching control-points from image J are found, producing an initial set of control point correspondences. To eliminate false correspondences we apply D. Lowe's ratio test [24] using his recommended ratio distance threshold of 0.75. Figure 4.10 shows control point correspondence between two sample image channels using SIFT control point extraction algorithm.

4.2.2 Graph Construction

We construct a complete undirected weighted graph $G = (C, E)$ (Step 3 in Figure 4.12) in which nodes C correspond to image channels in the dataset. Each undirected edge $[c_1, c_2]$ is annotated with the name of the control point extraction algorithm $\text{alg}(c_1, c_2)$ which produced the greatest average number of matches control point correspondences k when registering c_1 to c_2 or c_2 to c_1, and the weight of the edge $[c_1, c_2]$ is set to k. This approach allows for the possibility for different channel pairs to be registered using different control point extraction algorithms. Figure 4.11 (top) shows a sample undirected weighted graph $G = (C, E)$.

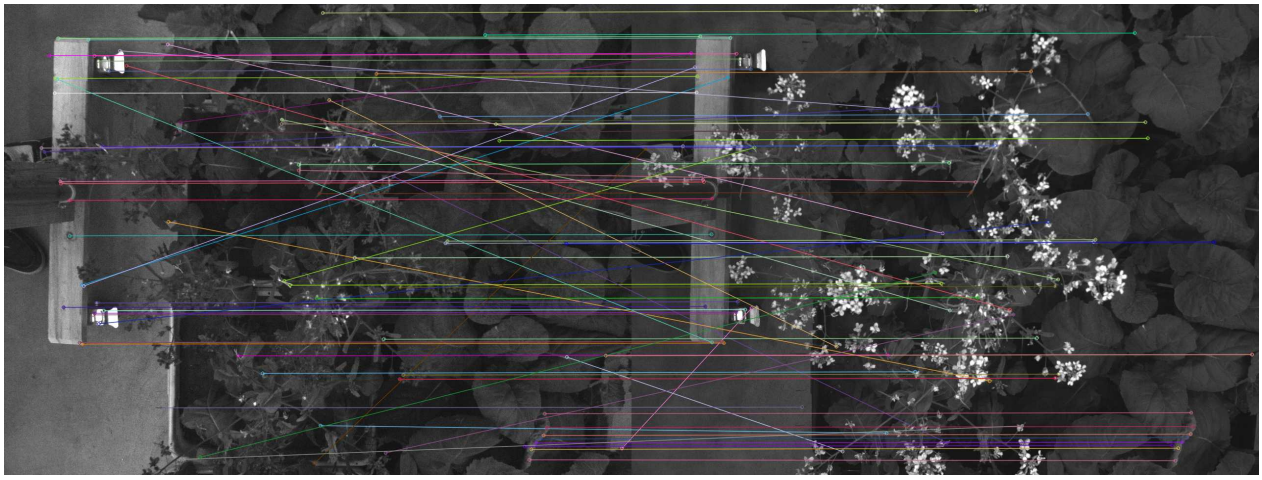

Figure 4.10: Control points correspondence match between two image channels.

4.2.3 Maximum Spanning Tree Construction

Kruskal's algorithm [20] is applied to G to determine its maximum spanning tree T (Step 4 in Figure 4.12). The edges in T represent the smallest set of edges in the graph that will allow all channels to be registered to some target channel while favoring channel pairs and control point extraction algorithms yielding larger average numbers of control-points. Figure 4.11 (right) shows a sample maximum spanning tree T constructed from undirected weighted graph G. The structure of T may or may not allow selection of a target channel to which all channels can be registered directly. This approach encourages the use of one or more intermediate channel registrations with larger numbers of average control-points over direct-to-target registrations with a small average number of control-points.

4.2.4 Finalizing the Registration Scheme

All that remains is to select the best target channel in T. We apply the Floyd-Warshall all-pairs-shortest-path algorithm [13] to T, treating, for this step, all edge weights as 1. The node with the smallest sum of distances from itself to all the other nodes is selected as the target channel for the registration scheme. This allows for the possibility of some channels being registered to intermediate channels instead of directly to the target when necessary but selects a target channel that minimizes the number of intermediate channel registrations. Then each undirected edge $[c_1, c_2]$ in T is converted into a directed edge (c_1, c_2) if c_2 is closer to the root, otherwise (c_2, c_1).

The output of the algorithm is a directed acyclic graph with a designated root (making it a tree). The root of the tree corresponds to the target channel to which all other images will be registered by following the directed edges. Each pair of channels for the edge (c_1, c_2) in the tree is registered using $\text{alg}(c_1, c_2)$. The output graph from the framework is the registration scheme for that dataset.

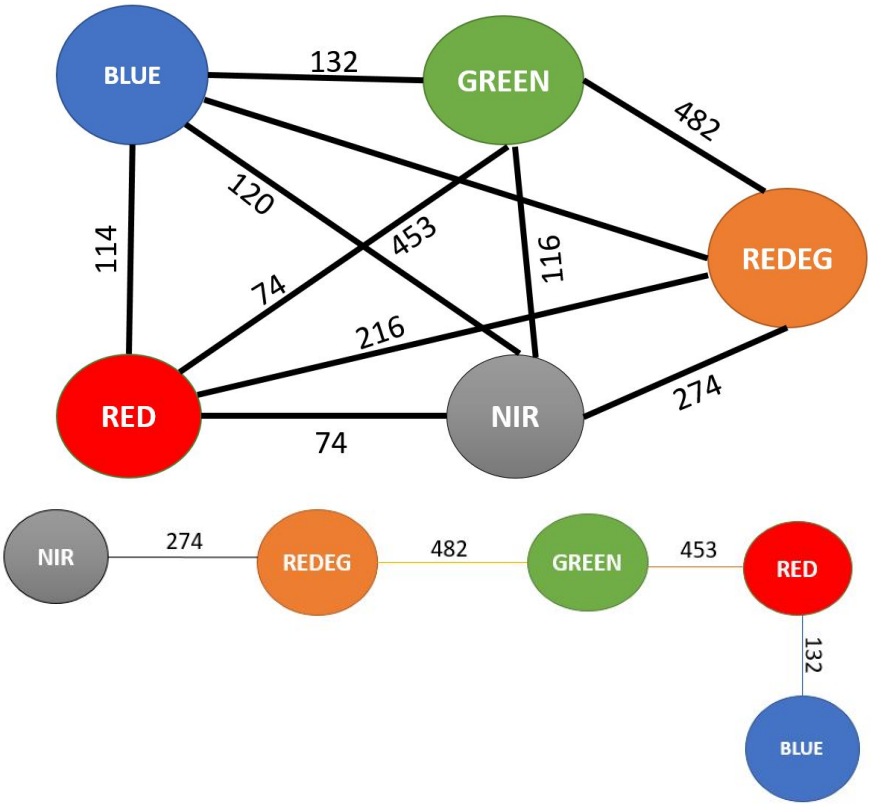

Figure 4.11: Graph construction (top) and maximum spanning tree diagram (bottom)

4.3 Registration Scheme To Image Transformation

Registration scheme generated from the data-driven multispectral image registration framework will be used to transform multispectral image channels and generate visual RGB images.

4.3.1 Image Transformation

Multi-spectral images will be registered directly-to-the-target channel or, through the intermediate channels following the directed acyclic graph of the data-driven registration scheme. We are using the registration scheme to transform images to generate registered multispectral image channels. Image transformation is happening following the image channel pairs (c_1, c_2) of the scheme using control point extraction algorithm $\text{alg}(c_1, c_2)$ which produced the greatest control point correspondence for that image channel pair (c_1, c_2). We are using control point correspondence, extracted from algorithms to calculate the homography matrix to scale the image channels c_2 according to the target image c_1. Warp-perspective function of OpenCV applies a perspective transformation to image channels c_2 based on the homography matrix generated from (c_1, c_2) [2]. Figure 4.13 shows the original image (left) and transformed registered image (right). Figure 4.14 is the

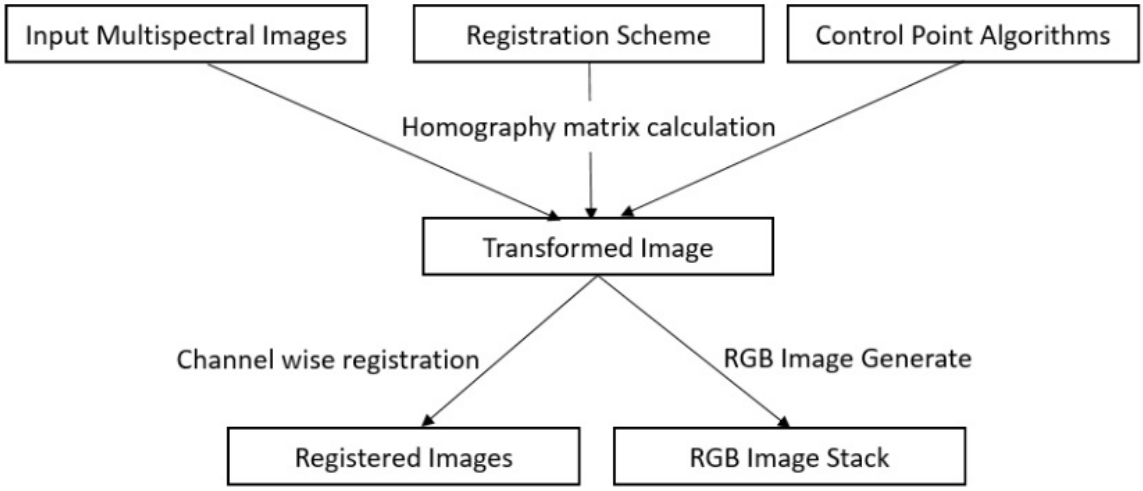

Figure 4.12: Image transformation process following data-driven registration scheme.

registered RGB sample image.

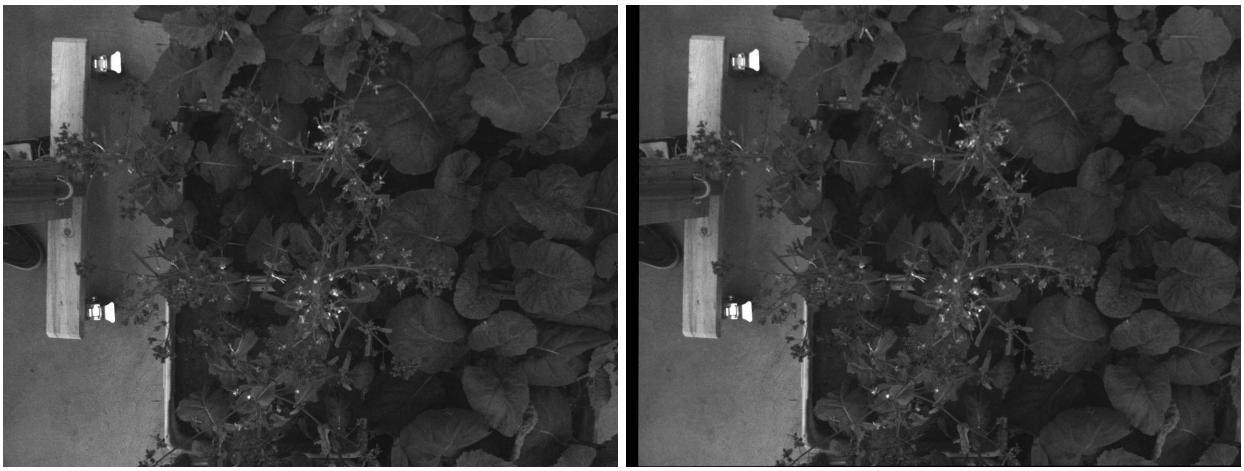

Figure 4.13: Image registration process following data-driven registration scheme. Left image is the raw single channel image and right image is the registered slightly shifted image

4.4 Framework Implementation

Data driven multi-spectral image registration framework is developed on python 3.2 programming language. It is supporting all the python versions from python 3.0. Integrated development environment that we have used is pycharm educational version. Control point extraction algorithms that we have used are SIFT, SURF, BRISK, ORB. All the control point extraction algorithms and geomatrical algorithms are separate from the main function in different functional code blocks. User is declaring the number of image channels, their directories, registered output channel directories and rgb output image directory in the main function of the

Figure 4.14: Registered RGB sample image of greenhouse data.

python program.

We have used OpenCV implementation of control point extraction algorithms in this framework. OpenCV is an open source image processing library available in different programming language. SURF, BRISK, ORB control point extraction algorithms are available in all the packages of OpenCV, SIFT is only available in OpenCV Contrib package. SIFT control point extraction algorithm is a patented algorithm by it's creator and free to use in academic and research purpose only. We have used OpenCV Contrib package where both patented and free control point extraction algorithms are available .

To measure the execution time of each control point extraction algorithm on both traditional and data-driven approach, we have used timeit module of python standard library. To calculate the execution time consistently, the code is running on a single thread using only one core of multi-core processor where, other processes are not affecting the runtime measurement process. During the process of independent control point finding and matching with other channels, we are initially loading all the channels of the same multi-spectral image. We have used CSV file as the file format to store raw data for further data analysis.

4.5 Limitations of the Framework

Traditional direct-to-target channel registration process takes one image as target channel and register all the other channels with respect to the target channel. In data-driven registration framework, we need to find out the average number of control point correspondences between all image channel pairs to generate registration scheme. Then, we follow the registration scheme to register image channels. Because of finding average control points of the entire dataset or, the partial dataset, data-driven image registration takes more

time than traditional approach.

Data-driven multispectral image registration framework is using four widely used control point extraction algorithms, SIFT, SURF, BRISK, ORB. Data-driven registration framework will be unable to generate registration scheme if these four control point extraction algorithms fail to perform in any dataset. Because of using SIFT control point extraction algorithm, this framework is only compatible with OpenCV contrib version.

4.6 Limitations of the Framework Evaluation

Multi-spectral image datasets that we have used in this research are either consists of four image channels or, five image channels. We were unable to find any publicly available dataset with more than five image channels. As data-driven multi-spectral image registration framework is dataset independent and should work on any number of image channels, we have tested this framework in only datasets consists of maximum five image channels.

Data driven multi-spectral image registration framework is generating the registration scheme using control point correspondences between image channel pairs and evaluating the performance using back projection error. Minimum number of control point correspondence required for image registration is four. If we get less the four control point correspondence between two image channel pairs, we are unable to generate homography matrix for that image channel pair and registration framework will be able to perform on that situation.

Chapter 5

Experiments and Results

Several experiments were designed with five different multispectral image dataset to evaluate and compare the performance of the data-driven image registration scheme and traditional direct-to-target channel registration scheme. In this chapter of the thesis, initially we will discuss the evaluation methodology and then we will briefly mention about dataset wise registration scheme generation process, registration scheme based analysis and performance statistics of different schemes.

5.1 Evaluation Methodology

We evaluated our registration framework on the five datasets described in Section 4.1 and used SIFT, SURF, BRISK, and ORB as the set of candidate control point extraction algorithms. We generated a registration scheme for each dataset, based on all images in the dataset. However, for large datasets, the registration scheme could be generated from a subset of images, and then applied all images in order to reduce computation time. In Section 6 we discuss some guidelines around when to use a subset of a dataset to generate the registration scheme.

We applied the generated registration scheme for each dataset and computed the average back-projection error (root-mean-square distance between a transformed control point and its corresponding control point) and overall number of control point correspondences for all channel-pair registrations performed by the scheme. Similarly, we computed the average runtime for an individual channel pair registration and the average number of control point correspondences used in an individual channel pair registration.

We compared our data-driven registration scheme to the traditional direct-to-target registration schemes, for which each edge is of the form (c, c_t) where c_t is the target channel, for each possible c_t. These represent all possible forms of the approach of choosing a target channel and registering each channel directly to the target.

5.2 Results and Analysis

We executed our evaluation methodology on all five datasets described in Section 4.1. We used SIFT, SURF, BRISK, and ORB as the set of candidate control point extraction algorithms. We have used entire dataset for

each of the experiment to generate registration scheme and made a comparison between registration scheme result and traditional approach.

5.2.1 Experiment 1: Canola Crop Breeding Data

Our first experiment is on the canola crop breeding multispectral drone image dataset. This dataset is consists of five image channels and they are blue, green, red, near-infrared and red-edge. The SIFT algorithm produced the greatest average number of control-points for every pair of channels. The average numbers of control-points from the algorithm are shown in Table 5.1. Each table entry is the greatest average number of control-points for registering the row channel to the column channel.

Image Channel	Blue	Green	Red	Near Infrared	Red Edge
Blue	–	370	1486	260	768
Green	352	–	4034	86	4606
Red	1461	3968	–	27	892
Near-Infrared	225	79	15	–	895
Red Edge	766	4755	949	942	–

Table 5.1: (Experiment 1) Greatest average number of control-points for each channel pair for the canola crop breeding dataset.

The resulting complete graph G is representing by table 5.1 of average numbers of control-points from SIFT control point extraction algorithm. Here, image channels are nodes in a graph and the average control point correspondence match between each image channel pair as the edges of the graph. Only the greatest average number of control-points match between image channel pairs are considered in the graph.

The output registration scheme after finding the minimum spanning tree of the graph in Figure 5.1, determination of root node, and conversion to a directed acyclic graph is shown in Figure 5.2. This registration scheme registers between channels most likely to produce large numbers of control-points, and indicates that the NIR and blue channels should be intermediately registered to the red-edge and red channels respectively.

Table 5.2 is the tabular representation of the registration scheme figure 5.2. It shows the step by step registration procedure for each image channel.

After applying the registration scheme in Figure 5.2 and all of the possible direct-to-target registration schemes in which no channel is registered to an intermediate channel, we compiled the results in Table 5.3.

Table 5.3 represents the overall results of both the traditional and data-driven registration approaches. The first six rows show results from schemes where there are no intermediate registrations and the row name is the target channel. The final row, labelled Our Reg Scheme, shows the results for the scheme generated by our algorithm. CP Matched is the average number of control-points found for an image pair registration

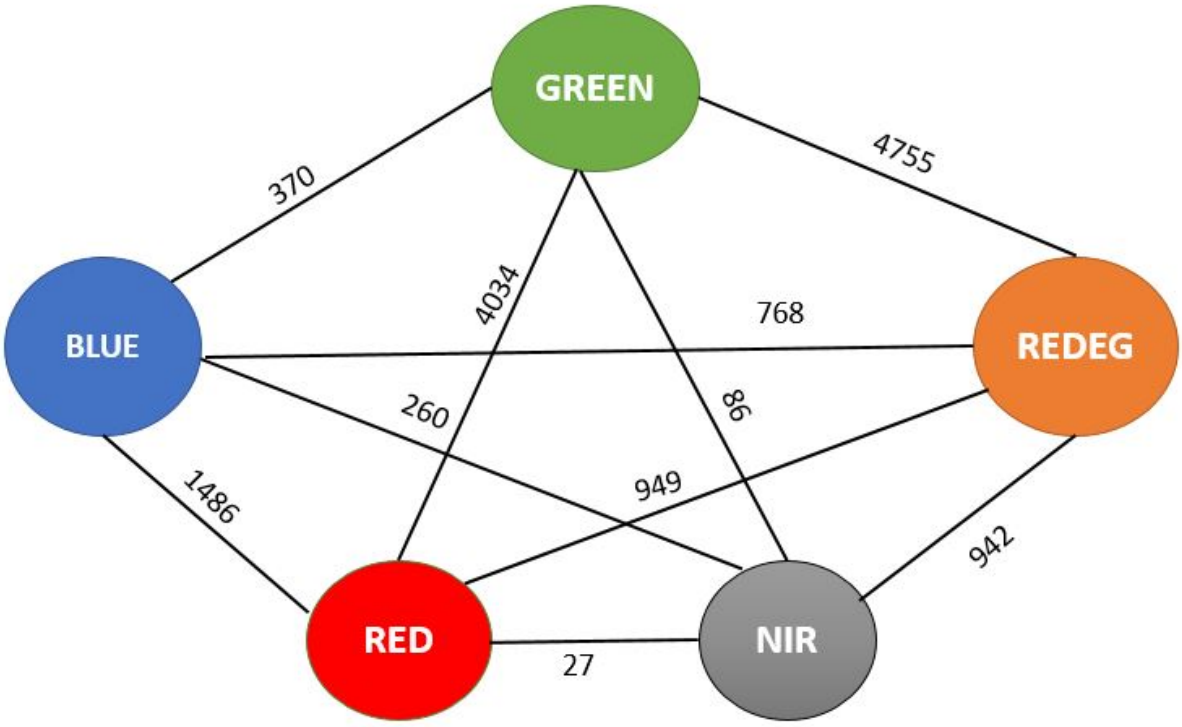

Figure 5.1: (Experiment 1) Graph construction from image channel pairs and control-points. For every edge $[c_1, c_2]$ in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$. This graph represents Table 5.1, where we have only considered the greatest average control-points between image channel pairs.

by the scheme. Runtime is the average runtime for an image pair registration by the scheme. BP error is the average back-projection error over all registrations performed by the scheme. All the numbers are mentioned along with their standard deviation values. The first five rows of the table are for the traditional single channel based registration approaches where these channels were target channel. Maximum average control-points correspondence of the table is 2804 from registration scheme, the minimum average back-projection error is 2.30 from the green channel based registration and minimum average run time is 0.35 seconds from the red-edge image channel based registration process.

Table 5.2 shows that the registration scheme generated by our algorithm resulted in more control-points being used on average in individual registrations compared to registering all channels directly-to-target channel for all possible targets, exceeding the result for the green channel by about 535 control-points.

Figure 5.3 is the histogram representation of back-projection error of both traditional direct-to-target channel registration and registration scheme. We have conducted KolmogorovSmirnov test in all the back-projection error representations and in all the tests, p-value was zero or, less than zero.

We are comparing the average runtime of direct-to-target channel and our registration scheme. We are not considering the registration scheme finding runtime for comparison with traditional approach. Runtime increases because of inappropriate control-point matching. Registration scheme is minimizing the chances of false control-points matches which is making the process faster than traditional approach. The average

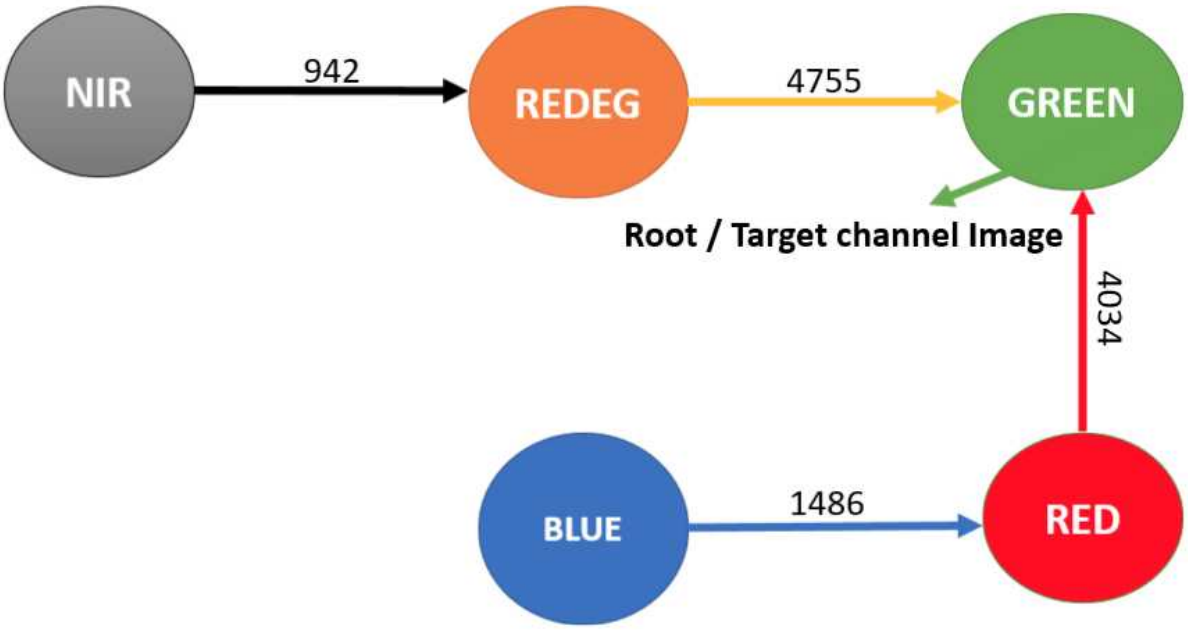

Figure 5.2: (Experiment 1) Registration scheme for generated by our algorithm for the Canola Crop UAV Dataset. For every edge (c_1, c_2) in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

runtime for our algorithm's scheme to register a pair of image channels was among the lowest, larger only than that for direct registration of all channels to near-infrared, but direct-to-near-infrared has the largest back-projection error, likely due to the small numbers of control-points that can be extracted. The average back-projection error for registration of a pair of image channels was lowest for our algorithm's registration scheme (Fig. 5.2). It was 87% of that of the best direct registration scheme. The standard deviation of back-projection error of our registration scheme shows that it is not only achieving lowest back-projection error, it is consistently showing less back-projection error across all image channels.

Images of canola crop breeding multispectral dataset are slightly misaligned from each other. If we stack the raw images together, we can see the misaligned visual RGB image from figure 5.4 left image. Figure 5.4 shows a sample result of a registration of RGB image channels registered with our registration scheme.

5.2.2 Experiment 2: Greenhouse Canola Plants Data

Our second experiment is on canola crop greenhouse multispectral image dataset. This dataset is consists of five image channels like the previous experiment and they are blue, green, red, near-infrared and red-edge. The SIFT algorithm produced the greatest average number of control-points for every pair of channels for this dataset as well. The average numbers of control-points from the algorithm are shown in Table 5.4. Each table entry is the greatest average number of control-points for registering the row channel to the column channel.

The resulting complete graph G is represented by table 5.4 of average numbers of control-points from

Channel	Registration Process
Green	Green
Red Edge	Red Edge - Green
Red	Red - Green
Near-Infrared	Near-Infrared - Red Edge - Green
Blue	Blue - Red - Green

Table 5.2: (Experiment 1) Registration scheme wise multispectral image registration table.

Channel	CP Matched	Run Time (s)	BP Error
Blue	721 ± 472	2.50 ± 1.74	0.45 ± 0.38
Green	2269 ± 719	2.30 ± 1.91	0.41 ± 0.18
Red	1589 ± 748	2.34 ± 1.68	0.45 ± 0.12
Red-Edge	1853 ± 552	2.34 ± 2.27	0.35 ± 0.42
Near-Infrared	303 ± 101	2.28 ± 1.42	0.75 ± 0.41
Our Reg Scheme	2804 ± 532	2.33 ± 0.51	0.30 ± 0.07

Table 5.3: (Experiment 1) Registration performance evaluation on the UAV canola crop dataset.

SIFT control point extraction algorithm. Here, image channels are nodes in a graph and the control point correspondence match between each image pairs as the edges of the graph. Only the greatest average number of control-points match between image channel pairs are considered in the graph.

The output registration scheme after finding the minimum spanning tree of the graph in Figure 5.5, determination of root node, and conversion to a directed acyclic graph is shown in Figure 5.6. This registration scheme registers between channels most likely to produce large numbers of control-points, and indicates that the near-infrared and blue channels should be intermediately registered to the red-edge and red channels respectively.

After applying the registration scheme in Figure 5.6 and all of the possible direct-to-target registration schemes in which no channel is registered to an intermediate channel, we compiled the results in Table 5.6. Table 5.6 represents the overall results of both the traditional and data-driven registration approaches. The first five rows of the table are for the traditional single channel based registration approaches where these channels were target channel. The last column of the table is showing the result of the data-driven registration scheme. Maximum average control-points correspondence of the table is 335 from registration scheme, the minimum average back-projection error is 3.05 from Red channel based registration and minimum average run time is 1.03 seconds from the blue channel based registration process.

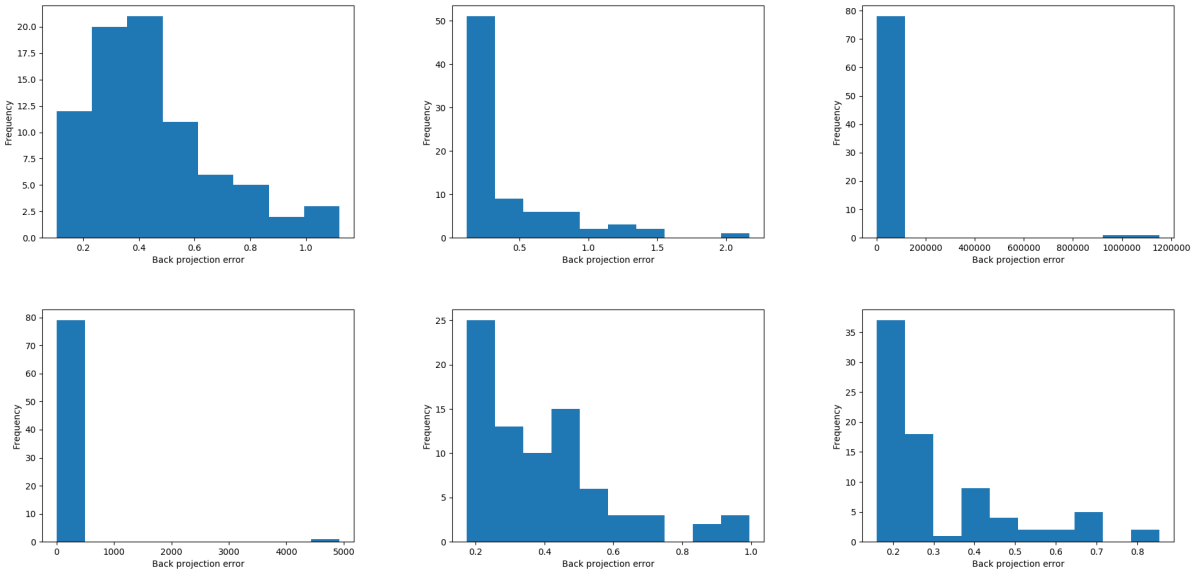

Figure 5.3: (Experiment 1) Histogram of Back projection error - Blue, Green, Red channel from top row, left to right and Near-Infrared, Red Edge and Registration Scheme from bottom row, left to right.

Compared with the 'direct-to-target' schemes, the registration scheme generated by our algorithm resulted in the greatest average number of control point correspondences between image channels exceeding the next closest result (Green) by 53. The difference between the smallest and largest average runtime for a single registration was only 0.12s. The scheme generated by our algorithm produced the second-smallest back-projection error and was almost the same as the back-projection error produced by the direct-to-red and direct-to-red-edge schemes. Direct-to-red channel based registration process showed lower back-projection error but, the standard deviation value of direct-to-red is higher than our registration scheme. It means, registration scheme consistently showed similar and lower back-projection error across all image channel

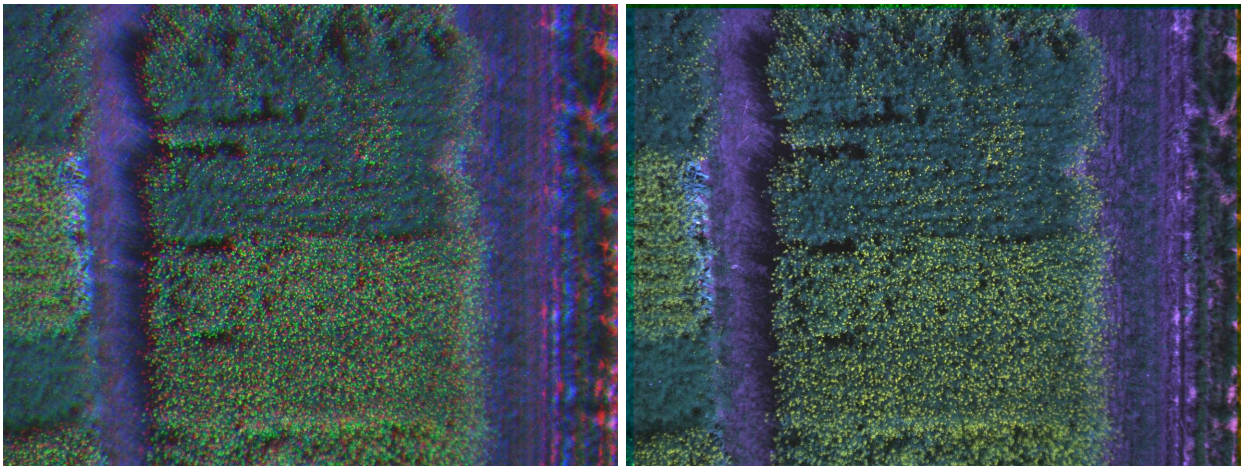

Figure 5.4: (Experiment 1) RGB image before (left) and after (right) registration

Image Channel	Blue	Green	Red	Near Infrared	Red Edge
Blue	–	101	100	35	90
Green	132	–	453	106	439
Red	114	425	–	57	185
Near-Infrared	74	116	74	–	274
Red Edge	120	482	216	269	–

Table 5.4: (Experiment 2) Greatest average number of contro-points for each channel pair for the canola crop greenhouse dataset.

Channel	Registration Process
Green	Green
Red Edge	Red Edge - Green
Red	Red - Green
Blue	Blue - Red - Green
Near-Infrared	Near-Infrared - Red Edge - Green

Table 5.5: (Experiment 2) Registration scheme wise multispectral image registration table.

pairs. Large back projection error of direct-to-blue channel and direct-to-near-infrared channel indicates wrong control-point correspondences between images channel pairs.

Figure 5.7 is the histogram representation of back-projection error of both traditional direct-to-target channel registration and registration scheme. We have conducted KolmogorovSmirnov test in all the back-projection error representations and in all the tests, p-value was zero or, less than zero.

Images of greenhouse multispectral images are misaligned with each other. If we stack the raw images together, we can see misaligned visual RGB greenhouse image from figure 5.8 left image. Figure 5.8 shows a sample result of an image registration with our registration scheme.

5.2.3 Experiment 3: UAV Mixed-Crop Dataset-1

Our third experiment is on UAV Mixed-Crop Dataset-1 captured with the SenseFly Sequoia camera. This dataset is consists of four image channels and they are green, red, near-infrared and red-edge. The SIFT algorithm produced the greatest average number of control-points for every pair of channels. The average numbers of control-points from the algorithm are shown in Table 5.7. Each table entry is the greatest average number of control-points for registering the row channel to the column channel.

The resulting complete graph G is representing the table 5.7 of average numbers of control-points from

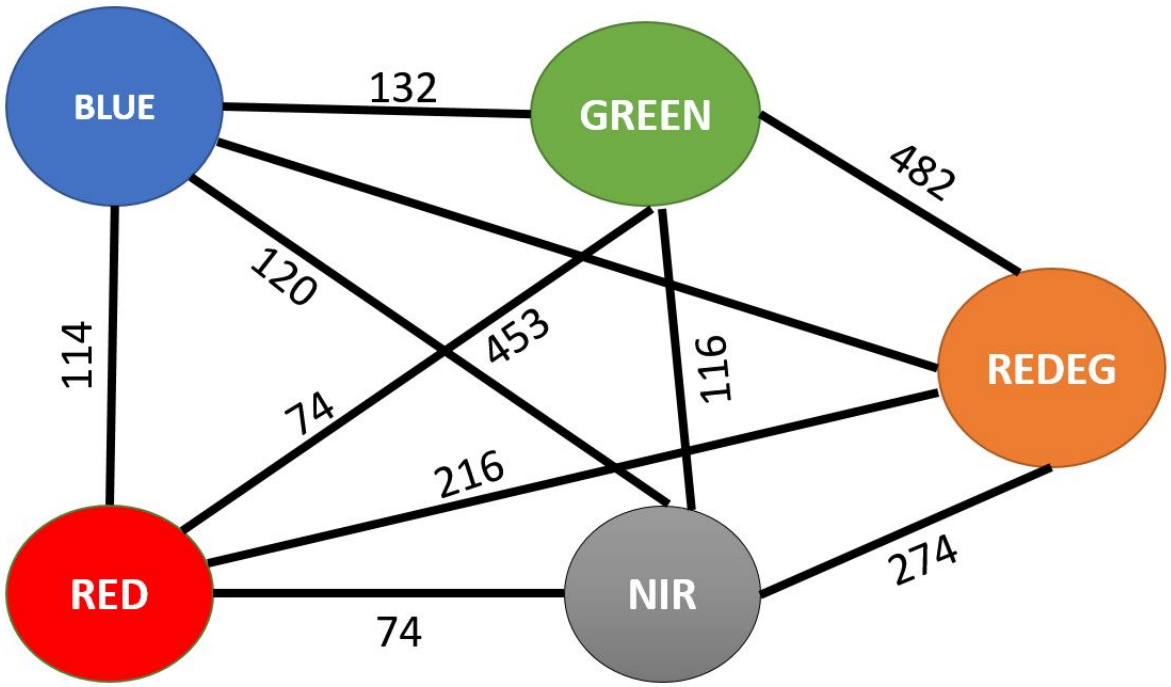

Figure 5.5: (Experiment 2) Graph construction from image channel pairs and control- points. For every edge $[c_1, c_2]$ in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

SIFT control point extraction algorithm. Here, image channels are nodes in a graph and the control point correspondence match between each image pairs as the edges of the graph. Only the greatest average number of control-points match between image channel pairs are considered in the graph.

The output registration scheme after finding the minimum spanning tree of the graph in Figure 5.9, determination of root node, and conversion to a directed acyclic graph is shown in Figure 5.10. This registration scheme registers between channels most likely to produce large numbers of control-points and indicates that Red channel should be intermediately registered to the green channel respectively.

Table 5.8 is the tabular representation of the registration scheme figure 5.10. It shows the step by step registration procedure for each image channels of UAV mixed crop dataset-1.

From table 5.8, we can see that, as the red-edge channel is our target channel. We are not registering the red-edge channel with other channels. Unlike other two previous experiments, here we have only 4 image channels. We have one intermediate registration step. Green channel and Near-infrared are registering with direct-to-target channel which is red-edge. For Red channel, we are going with an intermediate registration process. We are registering the red channel with respect to red-edge channel through the green channel.

Table 5.9 represents the overall results of both the traditional and data-driven registration approaches. The first four rows of the table are for the traditional single channel based registration approaches where these channels were target channel. The last column of the table is showing the result of the data-driven registration scheme. Maximum average control-points correspondence of the table is 2089 from registration

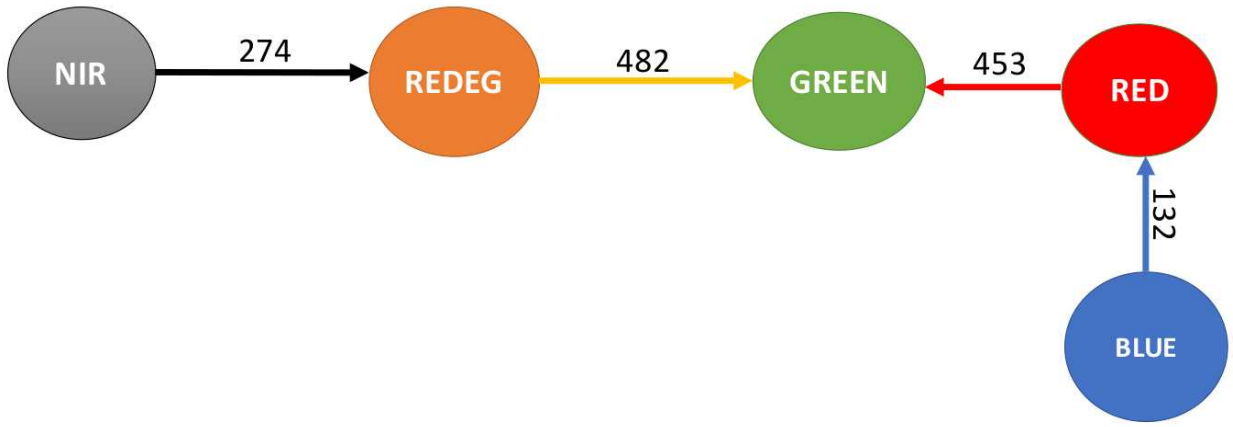

Figure 5.6: (Experiment 2) Registration scheme for generated by our algorithm for the Canola Crop UAV Dataset. For every edge (c_1, c_2) in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

Channel	CP Matched	Run Time (s)	BP Error
Blue	81 ± 7.14	1.03 ± 0.08	Failure
Green	282 ± 49.14	1.08 ± 0.15	4.05 ± 13.71
Red	195 ± 49.83	1.1 ± 0.11	3.05 ± 13.17
Red-Edge	271 ± 42.61	1.15 ± 0.15	3.1 ± 0.13
Near-Infrared	134 ± 40.43	1.18 ± 0.09	Failure
Our Reg Scheme	335 ± 51	1.15 ± 0.17	3.09 ± 0.13

Table 5.6: (Experiment 2) Registration performance evaluation on greenhouse dataset. The CP Matched, Run Time, and BP error are as in Table 5.2 along with their standard deviation values.

scheme, the minimum average back-projection error is 0.27 from the near-infrared channel based registration approach and minimum average run time is 1.60 seconds from the green channel based registration process.

Table 5.9 shows that the registration scheme generated by our algorithm resulted in more control-points being used on average in individual registrations compared to registering all channels directly to a target channel for all possible targets, exceeding the result for the green channel by about 333 control-points. The runtime for a single registration for our algorithm's scheme was comparable to most of the direct-to-target schemes with Green being noticeably smaller. Our algorithm's scheme resulted in a back-projection error slightly larger than direct-to-red-edge and direct-to-near-infrared but the back-projection errors were very low for all schemes (with the exception of direct-to-red). Standard deviation value of the back-projection error of our registration scheme shows that, the back-projection error of registration scheme was consistently low for all image channel pairs comparing to direct-to-near-infrared and direct-to-red-edge. Large back projection error of direct-to-red channel indicates wrong control-point correspondences between images channel pairs.

Figure 5.11 is the histogram representation of back-projection error of both traditional direct-to-target

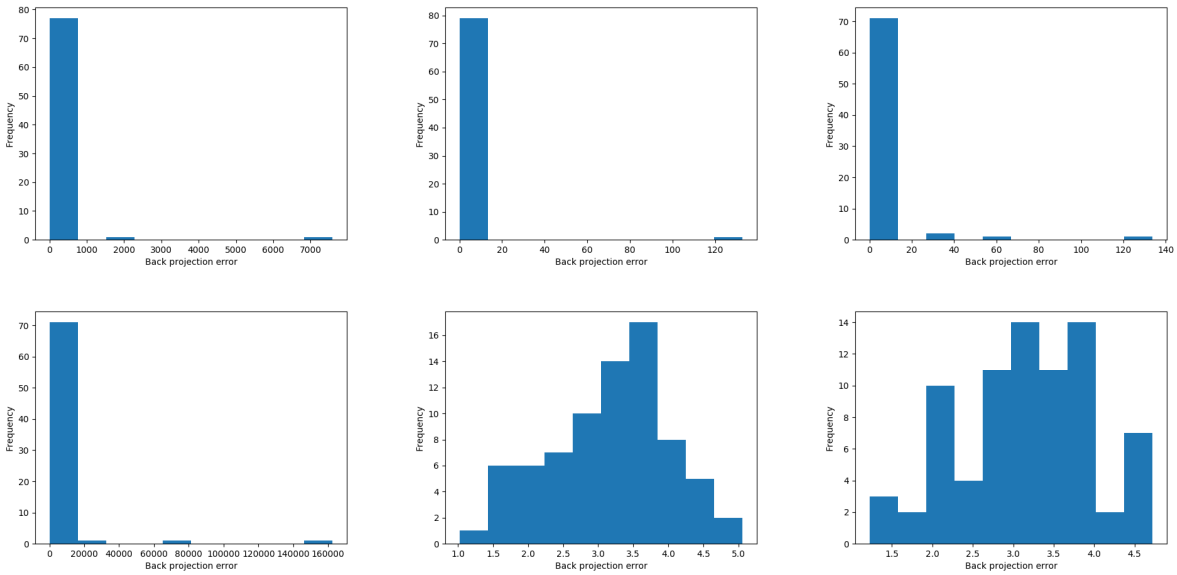

Figure 5.7: (Experiment 2) Histogram of Back projection error - Blue, Green, Red channel from top row, left to right and Near-Infrared, Red Edge and Registration Scheme from bottom row, left to right.

channel registration and registration scheme. We have conducted KolmogorovSmirnov test in all the back-projection error representations and in all the tests, p-value was zero or, less than zero.

Images of UAV mixed-crop dataset are slightly misaligned from each other. If we stack the raw images together, we can see the misaligned visual RGB image from figure 5.12 left the image. Figure 5.12 shows a sample result of a registration of RGB image channels registered with our registration scheme.

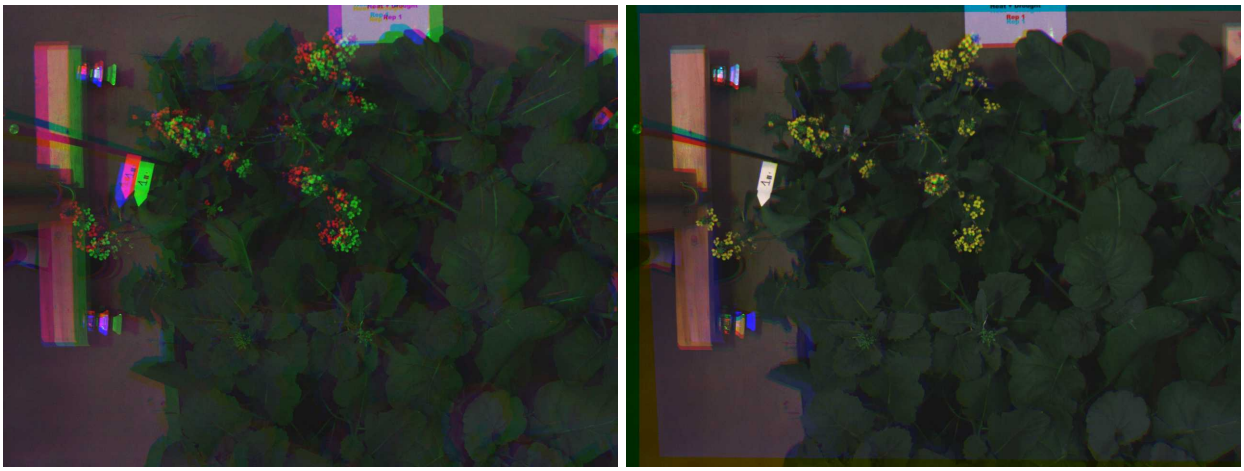

Figure 5.8: (Experiment 2) RGB greenhouse image before (left) and after (right) registration

Image Channel	Green	Near Infrared	Red	Red Edge
Green	–	257	1077	504
Near Infrared	251	–	74	4467
Red	1167	101	–	180
Red Edge	535	4565	169	–

Table 5.7: (Experiment 3) Greatest average number of control-points for each channel pair for the UAV Mixed-Crop Dataset-1.

Channel	Registration Process
Red-Edge	Red Edge
Green	Green - Red Edge
Near Infrared	Near Infrared - Red Edge
Red	Red - Green - Red-Edge

Table 5.8: (Experiment 3) Registration scheme wise multispectral image registration table.

5.2.4 Experiment 4: UAV Mixed-Crop Dataset-2

Our fourth experiment is on UAV mixed-crop dataset-2 captured with SenseFly eBee SQ drone carrying a SenseFly Sequoia camera. This dataset is consists of four image channels and they are green, red, near-infrared and red-edge. The SIFT algorithm produced the greatest average number of control- points for image channel pairs and SURF algorithms produced for rest of the three image channels pairs for this experiment. The average numbers of control-points from the algorithm are shown in Table 5.10. Each table entry is the greatest average number of control-points for registering the row channel to the column channel.

The resulting complete graph G is representing the table 5.10 of average numbers of control-points from different control point extraction algorithms such as SIFT and SURF for this experiment. Here, image channels are nodes in a graph and the control point correspondence match between each image pairs as the edges of the graph. Only the greatest average number of control-points match between image channel pairs are considered in the graph.

The output registration scheme after finding the minimum spanning tree of the graph in Figure 5.13, determination of root node, and conversion to a directed acyclic graph is shown in Figure 5.14. This registration scheme registers between channels most likely to produce large numbers of control-points and indicates that Red channel should be intermediately registered to the green channel respectively.

Table 5.11 is the tabular representation of the registration scheme figure 5.14. It shows the step by step

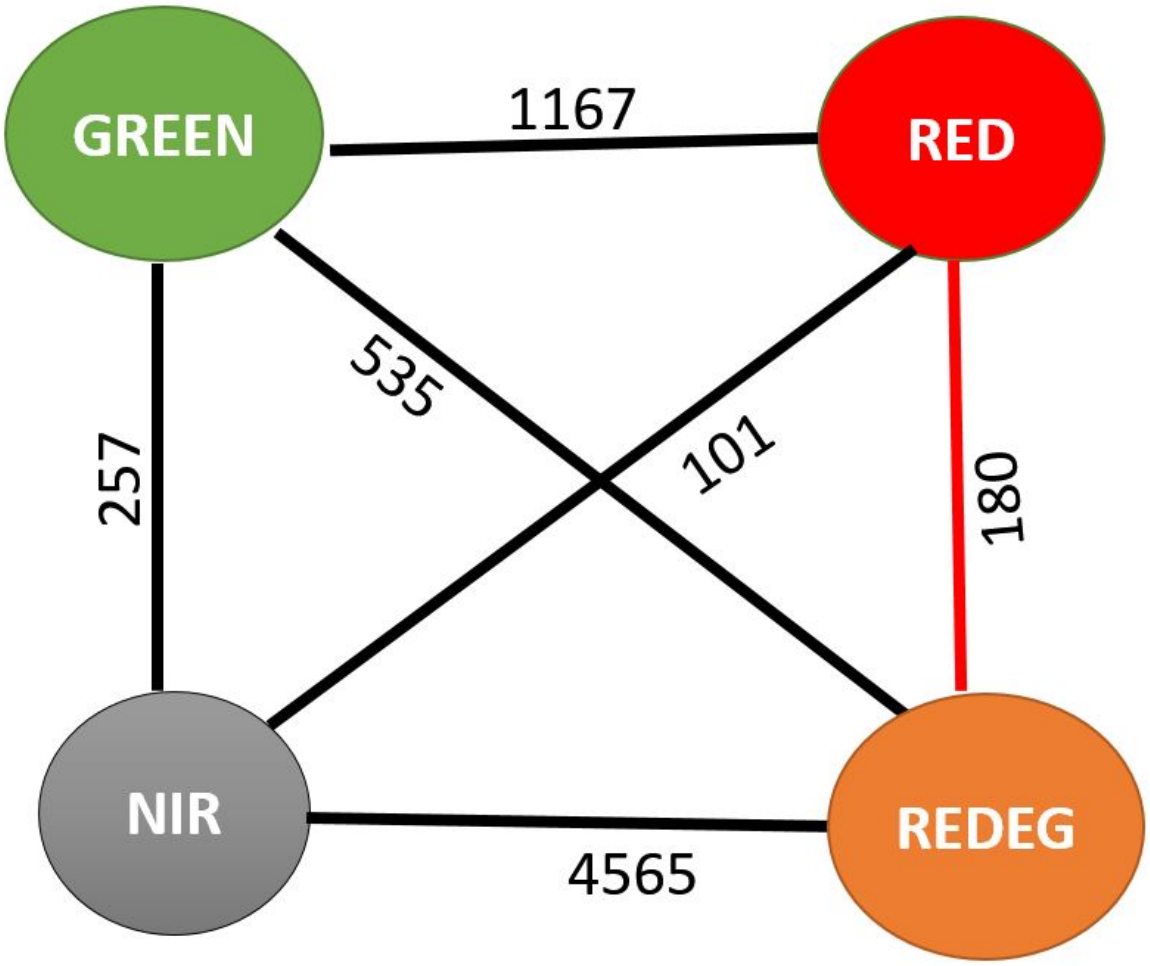

Figure 5.9: (Experiment 3) Graph construction from image channel pairs and control- points. For every edge $[c_1, c_2]$ in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

registration procedure for each image channels of UAV mixed-crop dataset-2.

From table 5.11, we can see that, as the red-edge channel is our target channel. We are not registering the red-edge channel with other image channels. Unlike first two previous experiments, here we have only 4 image channels. We are having one intermediate registration step. Green channel and Near-infrared are registering with direct-to-target channel which is red-edge. SURF control point extraction algorithm produced the greatest average control point correspondence between the green channel and red-edge channel registration and SIFT control point extraction algorithm between the near-infrared channel and red-edge channel. For Red channel, we are going with intermediate registration process using SURF control point extraction algorithm. We are registering the red channel with respect to red-edge channel through the green channel using SURF control point extraction algorithm in both the cases.

Table 5.12 represents the overall results of both the traditional and data-driven registration approaches.

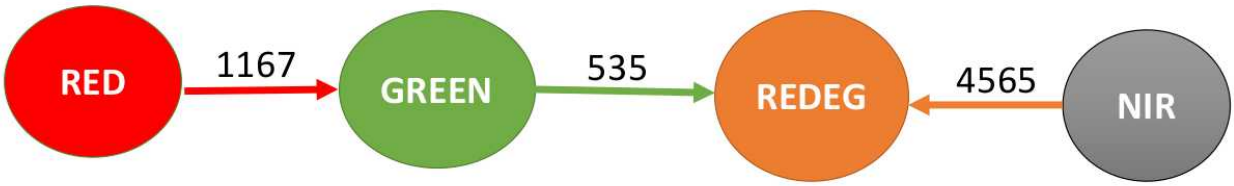

Figure 5.10: (Experiment 3) Registration scheme for generated by our algorithm for the UAV Mixed-Crop Dataset-1. For every edge (c_1, c_2) in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

Channel	CP Matched	Run Time (s)	BP Error
Green	612 ± 135.47	1.60 ± 0.41	0.37 ± 0.04
Red	482 ± 137.05	2.00 ± 1.28	Failure
Red-Edge	1756 ± 1462.78	2.10 ± 1.26	0.28 ± 0.06
Near-Infrared	1597 ± 1575.65	1.85 ± 0.90	0.27 ± 0.86
Our Reg Scheme	2089 ± 1426	1.93 ± 0.44	0.34 ± 0.01

Table 5.9: (Experiment 3) Registration performance evaluation on UAV mixed-crop dataset.

The first six rows show results from schemes where there are no intermediate registrations and the row name is the target channel. The final row, labelled Reg. Scheme, shows the results for the scheme generated by our algorithm. CP Matched is the average number of control-points found for an image pair registration by the scheme. Runtime is the average runtime for an image pair registration by the scheme. BP error is the average back-projection error over all registrations performed by the scheme. The first four rows of the table are for the traditional direct-to-target channel based registration approaches where these channels were target channel. The last column of the table is showing the result of the data-driven registration scheme. Maximum average control- points correspondence of the table is 1168 from registration scheme, the minimum average back-projection error is 0.45 from a green channel based registration approach and minimum average

Image Channel	Green	Near Infrared	Red	Red Edge
Green	–	182 (SURF)	661 (SURF)	451 (SURF)
Near Infrared	255 (SIFT)	–	151 (SIFT)	2318 (SIFT)
Red	729 (SURF)	192 (SIFT)	–	251 (SURF)
Red Edge	448 (SURF)	2325 (SIFT)	227 (SURF)	–

Table 5.10: (Experiment 4) Greatest average number of control-points for each channel pair for the UAV mixed-crop dataset-2.

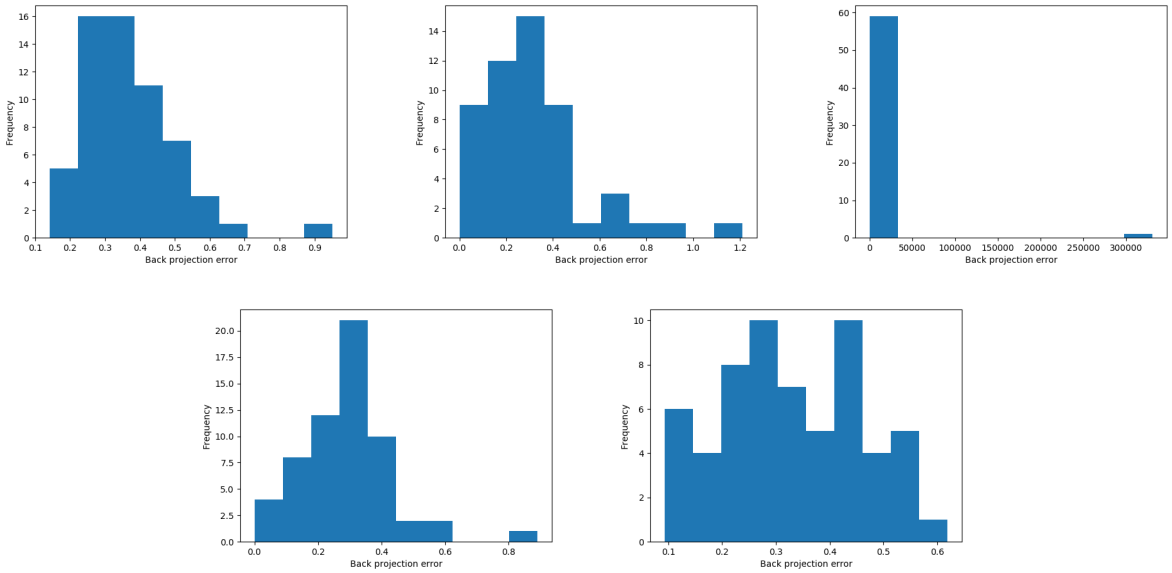

Figure 5.11: (Experiment 3) Histogram of Back projection error - Blue, Near-Infrared, Red channel from top row, left to right and Red Edge and Registration Scheme from bottom row, left to right.

run time is 1.45 seconds from the green channel based registration process.

Table 5.12 shows that the registration scheme generated by our algorithm resulted in more control-points being used on average in individual registrations compared to registering all channels directly to a target channel for all possible targets, exceeding the result for the near-infrared channel by about 168 control-points. Direct-to-green channel based registration approach showed less runtime and back-projection error than our registration scheme. Registration scheme showed second lowest back=projection error. Standard deviation value of the back-projection error of our registration scheme shows inconsistent back-projection error value for all image channel pairs in this experiment. Large back projection error of direct-to-red channel and

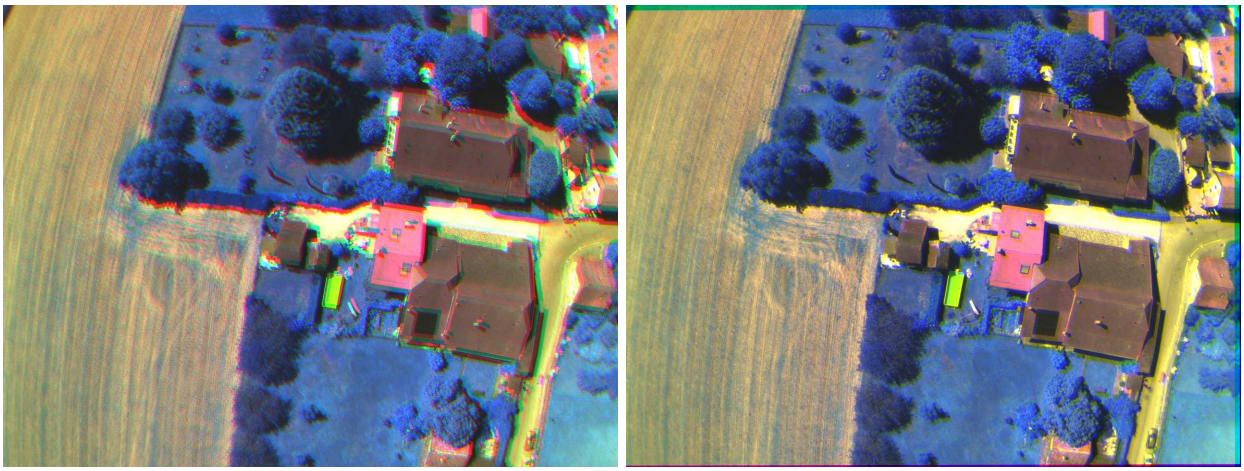

Figure 5.12: (Experiment 3) RGB image before (left) and after (right) registration

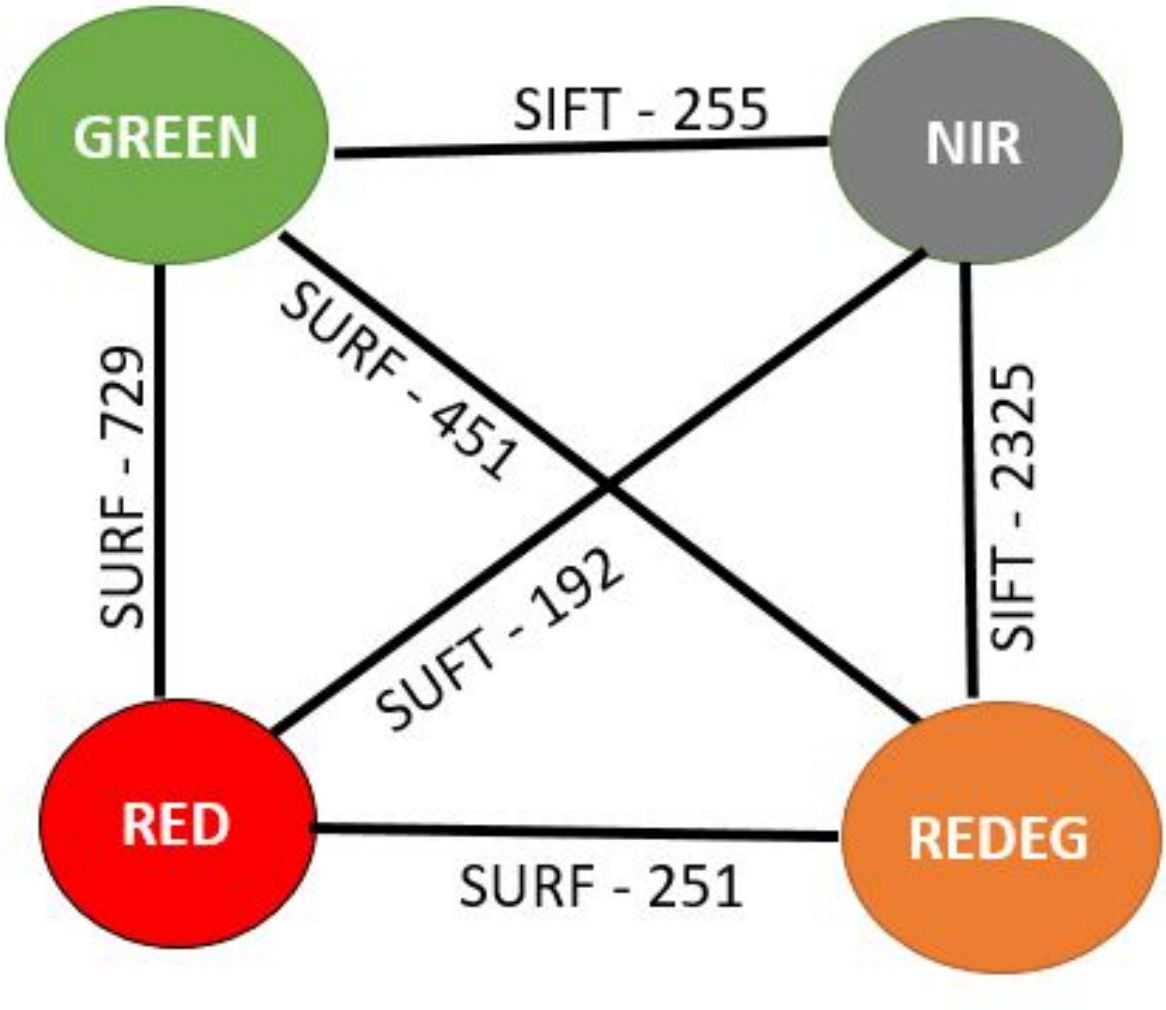

Figure 5.13: (Experiment 4) Graph construction from image channel pairs and control- points. For every edge $[c_1, c_2]$ in the graph, alg(c_1, c_2) = SIFT&SURF mentioned accordingly.

direct-to-red-edge channel indicates wrong control-point correspondences between images channel pairs.

Figure 5.15 is the histogram representation of back-projection error of both traditional direct-to-target channel registration and registration scheme. We have conducted KolmogorovSmirnov test in all the back-projection error representations and in all the tests, p-value was zero or, less than zero.

Images of UAV mixed-crop dataset-2 are slightly misaligned from each other. If we stack the raw images together, we can see the misaligned visual RGB image from figure 5.16 left image. Figure 5.16 shows a sample result of a registration of RGB image channels registered with our registration scheme.

5.2.5 Experiment 5: TokyoTech Dataset

Our fifth and last experiment is on TokyoTech multispectral dataset.This dataset is consists of five image channels and they are blue, green, red, Cy and Or. The SIFT algorithm produced the greatest average number

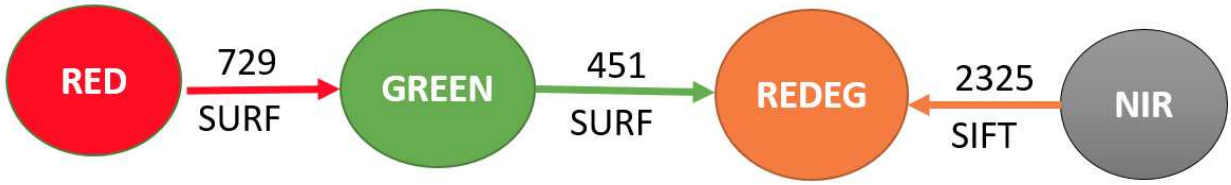

Figure 5.14: (Experiment 4) Registration scheme for generated by our algorithm for the UAV mixed-crop dataset-2. For every edge (c_1, c_2) in the graph, alg(c_1, c_2) = SIFT&SURF mentioned accordingly.

Channel	Registration Process
Red Edge	Red-Edge
Green	Green - Red-Edge
Near Infrared	Near Infrared - Red Edge
Red	Red - Green - Red-Edge

Table 5.11: (Experiment 4) Registration scheme wise multispectral image registration table.

of control- points for every pair of channels. The average numbers of control-points from the algorithm are shown in Table 5.13.

The resulting complete graph G is representing the table 5.13 of average numbers of control-points from SIFT control point extraction algorithm. Here, image channels are nodes in a graph and the control point correspondence match between each image pairs as the edges of the graph. Only the greatest average number of control-points match between image channel pairs are considered in the graph.

The output registration scheme after finding the minimum spanning tree of the graph in Figure 5.17, determination of root node, and conversion to a directed acyclic graph is shown in Figure 5.18. This registration scheme registers between channels most likely to produce large numbers of control-points and indicates that the blue and red channels should be intermediately registered to the cy and or channels respectively.

Table 5.14 is the tabular representation of the registration scheme figure 5.18. It shows the step by step registration procedure for each image channels of TokyoTech multispectral dataset.

From table 5.14, we can see that, as the green channel is our target channel. We are registering or channel and cy channel with direct-to-target channel which is green channel. For blue channel and red channel, we are going with intermediate registration process like previous experiments.

Table 5.15 represents the overall results of both the traditional and data-driven registration approaches. The first five rows of the table are for the traditional single channel based registration approaches where these channels were target channel. The last column of the table is showing the result of the data-driven registration scheme. Maximum average control-points correspondence of the table is 2483 from registration scheme, the minimum average back-projection error is 0.3362 from our registration scheme and minimum

Channel	CP Matched	Run Time(s)	BP Error
Green	431 ± 116.06	1.45 ± 0.17	0.45 ± 0.05
Near-Infrared	908 ± 874	2.26 ± 0.77	Failure
Red	390 ± 110	2.34 ± 0.62	Failure
Red Edge	1000 ± 861	1.70 ± 0.77	Failure
Reg Scheme	1168 ± 764	1.79 ± 0.81	8.21 ± 62.19

Table 5.12: (Experiment 4) Registration performance evaluation on UAV mixed-crop dataset-2.

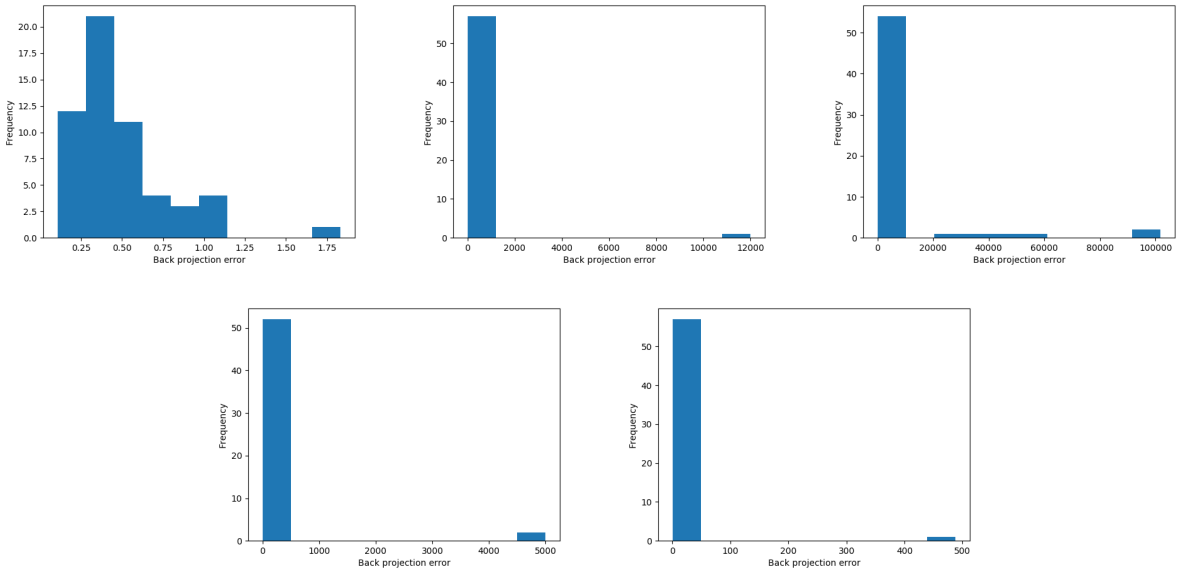

Figure 5.15: (Experiment 4) Histogram of Back projection error - Blue, Near-Infrared, Red channel from top row, left to right and Red Edge and Registration Scheme from bottom row, left to right.

average run time is 1.80 seconds from the red channel based registration process.

Table 5.15 shows that the registration scheme generated by our algorithm resulted in more control-points being used on average in individual registrations compared to registering all channels direct-to-target channel for all possible targets, exceeding the result for the green channel by about 639 control-points. The runtime for a single registration for our algorithm's scheme was comparable to most of the direct-to-target schemes with red being noticeably smaller. Our algorithm's scheme resulted in consistently lower back-projection error than all the other single channel based traditional approaches.

Figure 5.19 is the histogram representation of back-projection error of both traditional direct-to-target channel registration and registration scheme. We have conducted KolmogorovSmirnov test in all the back-projection error representations and in all the tests, p-value was zero or, less than zero.

Images of TokyoTech multispectral dataset are slightly misaligned after rotating them in a random order

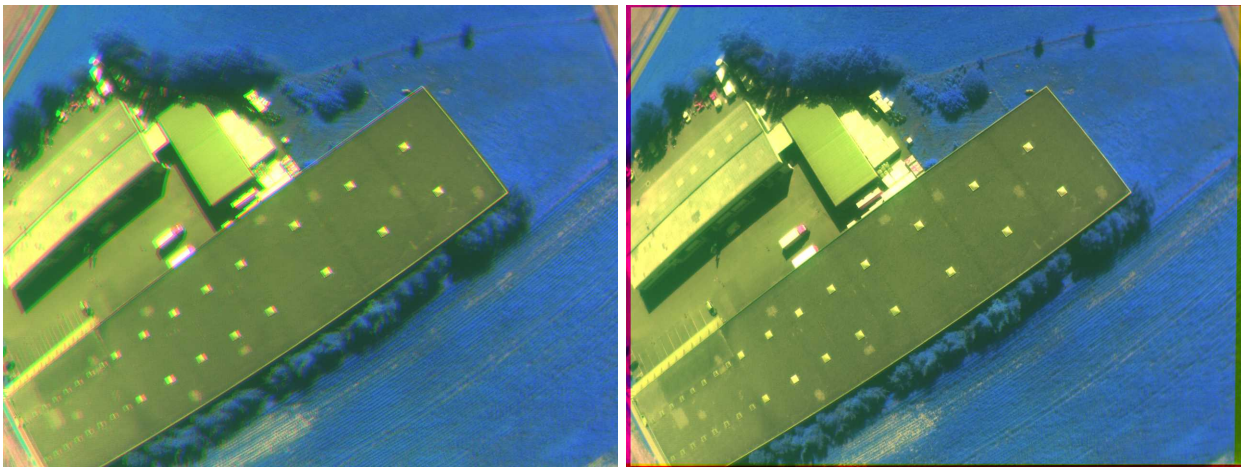

Figure 5.16: (Experiment 4) RGB image before and after (right) registration

Image Channel	Blue	Cy	Green	Or	Red
Blue	–	3052	1479	946	741
Cy	3066	–	1820	1106	880
Green	1552	1907	–	2416	1501
Or	947	1089	2329	–	2788
Red	718	847	1400	2732	–

Table 5.13: (Experiment 5) Greatest average number of control-points for each channel pair for the TokyoTech multispectral dataset.

from -3 degree to +3 degree. If we stack the rotated images together, we can see the misaligned visual RGB image from figure 5.20 left the image. Figure 5.20 shows a sample result of a registration of RGB image channels registered with our registration scheme.

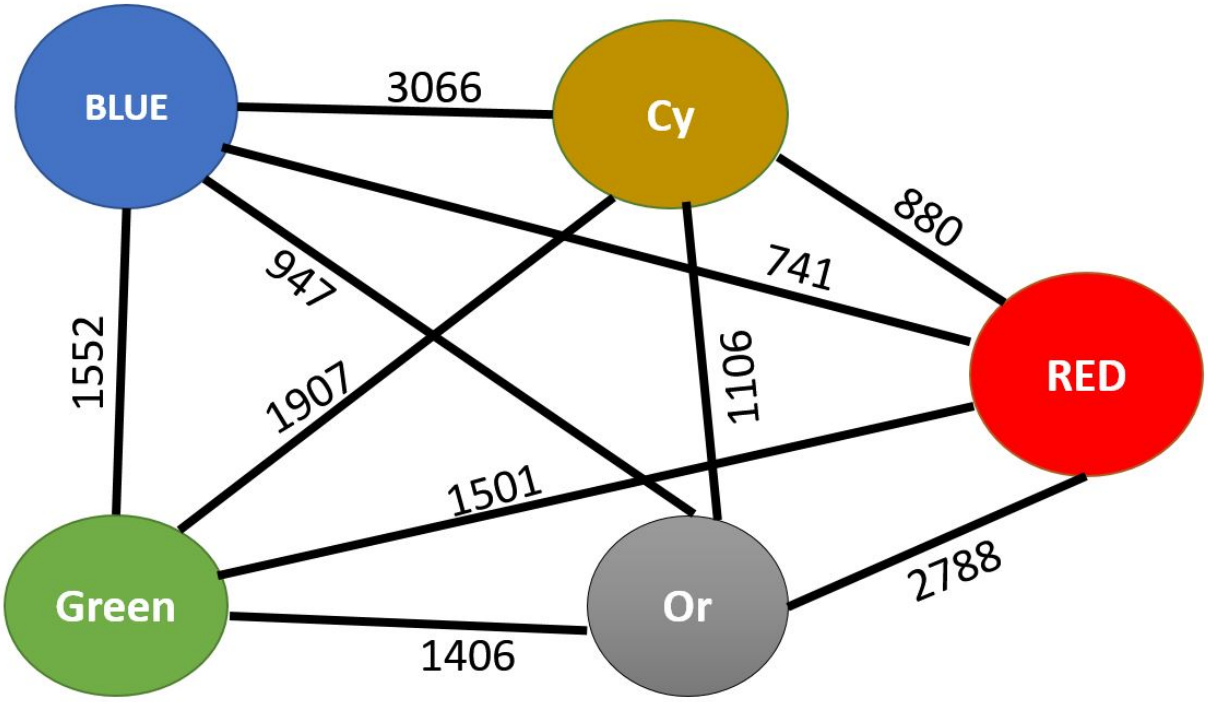

Figure 5.17: (Experiment 5) Graph construction from image channel pairs and control- points. For every edge $[c_1, c_2]$ in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

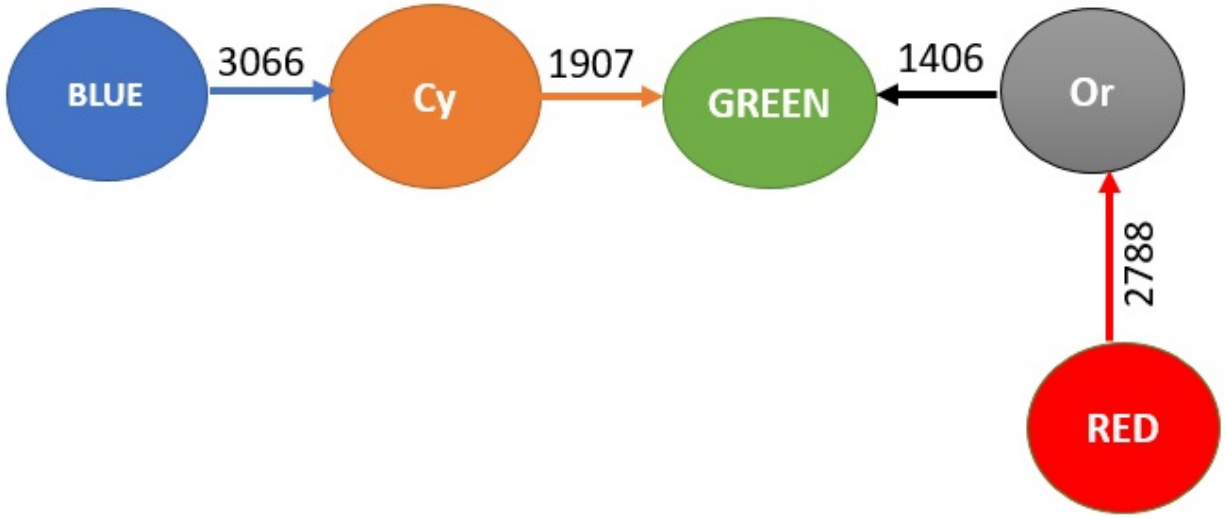

Figure 5.18: (Experiment 5) Registration scheme for generated by our algorithm for the TokyoTech multispectral dataset. For every edge (c_1, c_2) in the graph, $\text{alg}(c_1, c_2) = \text{SIFT}$.

Channel	Registration Process
Green	Green
Cy	Cy - Green
Red	Red - Or - Green
Blue	Blue - Cy - Green
Or	Or - Green

Table 5.14: (Experiment 5) Registration scheme wise multispectral image registration table.

Channel	CP Matched	Run Time (s)	BP Error
Blue	1554 ± 673	1.85 ± 0.43	0.48 ± 0.06
Cy	1718 ± 679.60	1.9 ± 0.59	0.48 ± 0.13
Green	1844 ± 388.73	2 ± 0.41	0.48 ± 0.07
Or	1788 ± 488	1.9 ± 0.55	0.42 ± 0.09
Red	1424 ± 455.65	1.8 ± 0.31	0.47 ± 0.26
Registration Scheme	2483 ± 275.46	2.18 ± 0.70	0.33 ± 0.08

Table 5.15: (Experiment 5) Registration performance evaluation on TokyoTech multispectral dataset.

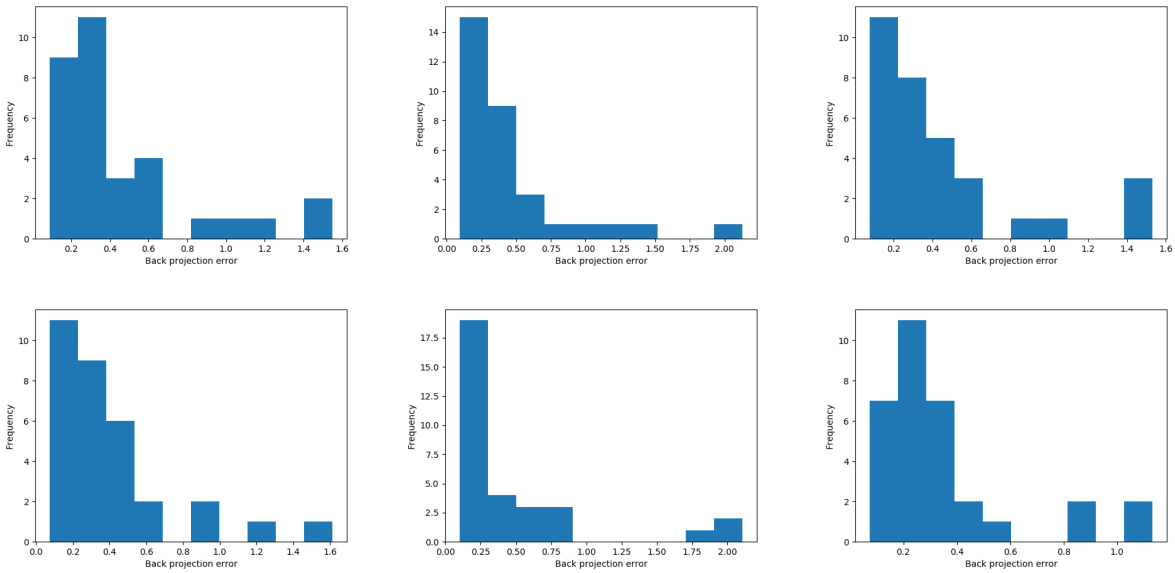

Figure 5.19: (Experiment 5) Histogram of Back projection error - Blue, Cy, Green channel from top row, left to right and Or, Red and Registration Scheme from bottom row, left to right.

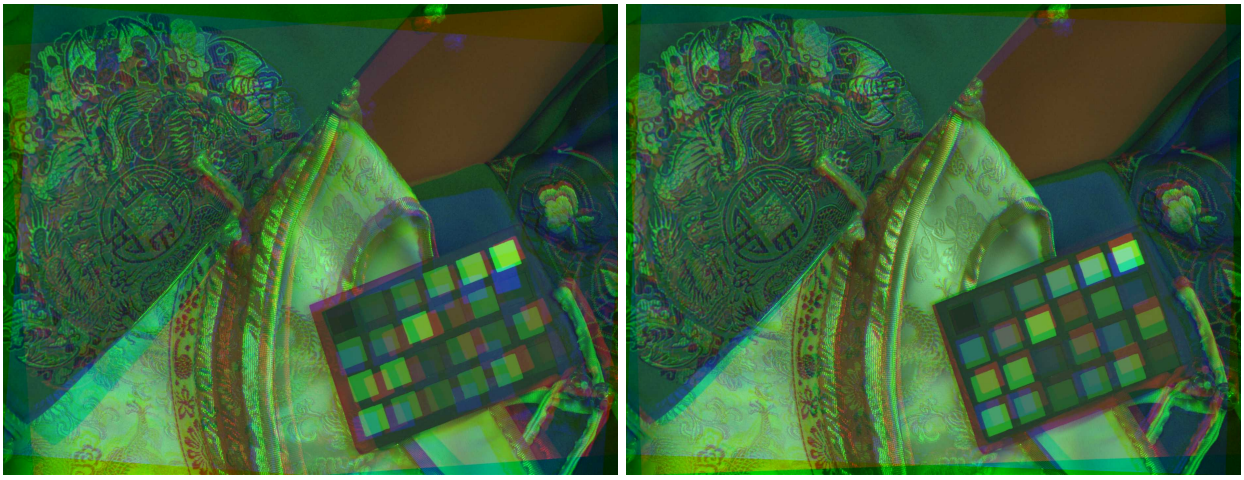

Figure 5.20: (Experiment 5) RGB image before (left) and after (right) registration

Chapter 6

Discussion

6.1 Discussion and Analysis

Across all five experiments, our registration scheme frequently produced more control point correspondence matches than traditional direct-to-target-channel registration approach and commonly resulted in a lower back-projection error. In experiment 1, the data-driven registration scheme produced maximum control point correspondences with minimum back-projection error consistently across all image channel pairs than traditional direct-to-target-channel registration approaches. In experiment 2, the direct-to-red scheme showed less back-projection error than the registration scheme. The standard deviation of back-projection error of registration scheme is 0.1332, and red channel-based approach is 13.171 from table 5.6. This indicates that there were likely less catastrophic failure cases using our scheme. In experiment 3, the direct-to-red-edge and direct-to-near-infrared schemes exhibited less back-projection error than our algorithm's registration scheme. The standard deviation of mean back-projection error for our algorithm's scheme was much less than all of the direct-to-target schemes. This suggests that our algorithm's schemes are registering with more consistent back-projection error and fewer catastrophic failures. Experiment 4 is the only experiment where our registration scheme showed inconsistency and high back-projection error from table 5.12. The Direct-to-green scheme approach showed consistently less back-projection error in this experiment than registration scheme but, registration scheme showed 737 more control point correspondences than direct-to-green. In experiment 4, the registration for direct-to-near-infrared channel showed high back-projection error with respect to all the image channels. Because of the high back-projection error of the direct-to-near-infrared channel, average back-projection error of registration scheme became very high in this experiment. In experiment 5, our registration scheme consistently showed lower back-projection error than all the traditional approaches (table 5.15). Experiment 4 on UAV mixed-crop dataset2 showed overall very high back-projection error comparing to all the other experiments. Though ratio test [24] has discarded 90% of the false control point matches, for this dataset control-points that matched well showed high back-projection error. Both direct-to-target channel based registration and data-driven registration scheme showed a similar result. Other than experiment 4, the SIFT control point extraction algorithm showed superiority in all of our experiments. In experiment 4, both SIFT and SURF provided the greatest average control point correspondence between image channel pairs. The four out of five of our experiments, SIFT showed superiority where only in one experiment SURF

provided maximum average control point correspondence for three image pairs and SIFT got for other three pairs. We have used four different control point algorithms in this framework and if we have used only SIFT, there would be cases like experiment 4, where we would have obtained bad result according to registration scheme. The reason behind using four different control point extraction algorithms is, different control point extraction algorithms will perform well in different scenarios and datasets.

We have conducted KolmogorovSmirnov test in all the histogram representations of the back-projection error. P-value of all the KolmogorovSmirnov tests were zero or, less than zero. Back-project error across all the experiments are not normally distributed. The data-driven image registration framework is considering only the average control point correspondences to generate registration scheme. Random distribution of the back-projection error across all image channel pairs is the reason for not considering back-projection error to generate registration scheme. Figure 5.3, 5.7, 5.11, 5.15, 5.19 shows that, other than experiment 4, data-driven registration scheme has consistently given less back-projection error and avoided catastrophic failures successfully. Registration scheme is consists of image channel pairs with maximum average control point matches. From the histogram representations of the back-projection error, it is visible that poor number of control point matches between image channel pairs tends to have catastrophic failures. By considering maximum average control point matches, registration framework has successfully avoided failure situations.

6.2 The apparent superiority of SIFT

In four out of five experiments, SIFT universally resulted in more control- points extracted than the other control point extraction algorithms. We investigated this further using the canola crop breeding dataset. For all the images in the dataset, every channel was directly registered to every other channel, and the average number of control-points found when registering to a particular target channel was computed for each of the control point extraction algorithms. These results are shown in Figure 6.1

The average number of control-points extracted consistently highest for SIFT. The average number of control-points extracted by ORB on this dataset was very small and was zero for the blue channel. The average number of control-points extracted by SURF was the second best among all the four control point extraction algorithms. Direct-to-near-infrared channel registration produced the least average number of control-points amongst all the direct-to-target channel registration approaches.

In the same way, we computed the average back-projection error for all control point extraction algorithms. Figure 6.2 shows the average back-projection error is generally best for SIFT, though BRISK is a strong competitor and had a lower back-projection error than SIFT for registrations to the red channel. There is no bar for the blue channel for ORB because there were no control point correspondences found for the blue channel by ORB.

In the same way, we computed the average runtime for all control point extraction algorithms as well. Figure 6.3 shows the average runtime is best for SURF, though BRISK is a strong competitor and showed

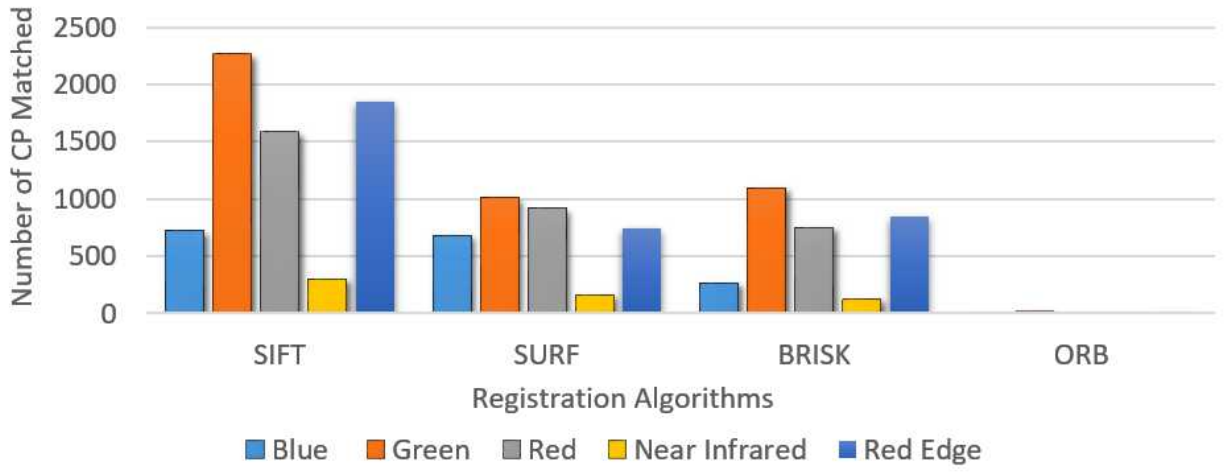

Figure 6.1: Average number of control-points matched for each control point extraction algorithm.

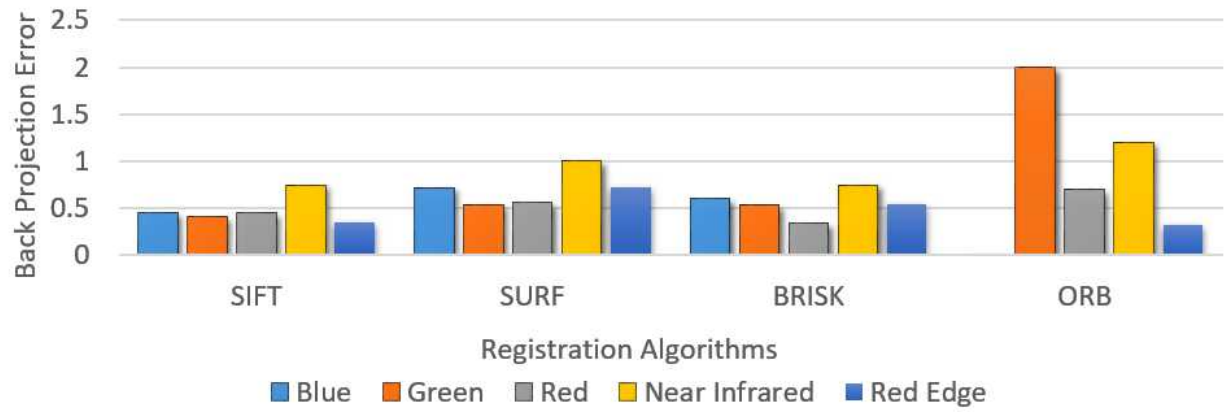

Figure 6.2: Average back-projection error for target channels in the canola crop UAV dataset.

second best runtime among all the four control point extraction algorithms. SIFT is the slowest control point extraction algorithm, and ORB is the fastest control point extraction algorithm but, SIFT produced the most number of average control point correspondences with lower back-projection error and ORB produced the least number of control point correspondences with the highest back-projection error.

While there is evidence that SIFT is frequently superior, there is also evidence that strong competitors such as BRISK may result in better performance for registrations between some channels in some datasets justify the idea of allow different control point extraction algorithms for registration of various misaligned channel pairs in a registration scheme can provide superior performance.

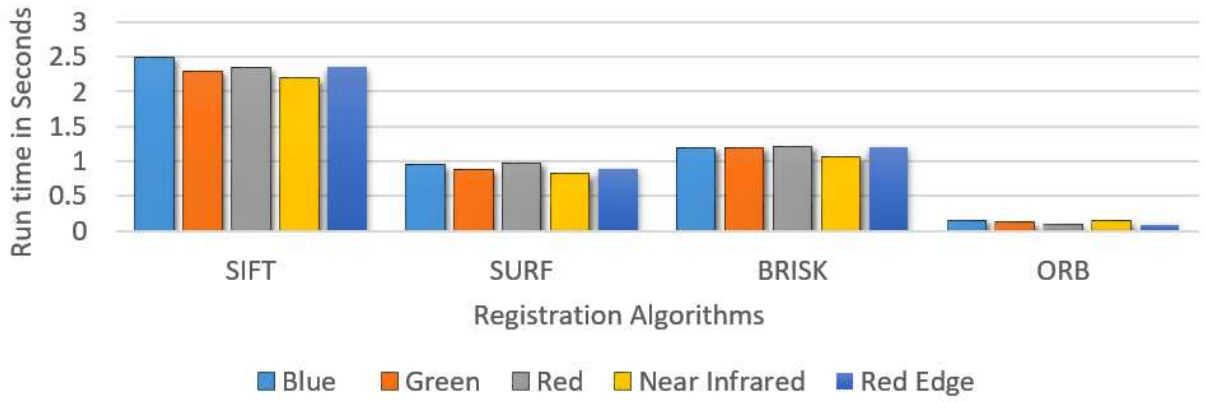

Figure 6.3: Average runtime for target channels in the canola crop breeding dataset.

6.3 Generating Registration Schemes using Partial Datasets

We call a dataset homogeneous when all the content of the dataset are similar to each other. In all of our datasets, we have the same number of image channels with similar contents. The datasets we used in this work were not so large as to prohibit the use of the entire dataset to determine the best registration scheme. However, if the dataset is homogeneous in that it consists of images of the same type of scene captured in the same way using the same equipment, then it is not necessary to use the entire dataset to generate the registration scheme. To verify that, we generated registration schemes using only one-fifth of each of our five test datasets. Registration schemes generated using one-fifth images of the entire dataset showed similar results of the schemes generated using the whole test dataset.

The resulting registration schemes generally produced results comparable to using the entire datasets. On average it takes about 360 seconds to perform all of the channel registrations using each of the control point extraction algorithms for one five-channel image of 1280×960 pixels. The processing time to generate a registration scheme for a dataset (or subset) thus depends on the number of images, number of image channels, the image size in pixels, and the hardware configuration. We have used a PC with an Intel Core i7-2600 3.40 GHz processor with 8 cores and 16 Gb of memory to generate evaluation results.

Chapter 7

Conclusion

A new data-driven image registration framework was proposed in this thesis. The traditional approach for multispectral image registration process is to select a target channel and register all other image channels to the target. There is no objective evidence-based method to select a target. The registration schemes generated by our algorithm frequently performed better in our experiments than the traditional direct-to-target approach and provided an objective evidence-based process of choosing a target channel. The advantages of our algorithm include its data-driven approach that can generate customized registration schemes for specific multispectral image datasets, the allowance for channels to be indirectly registered to the target channel by registering first to one or more intermediate channels potentially avoiding registration between channel pairs where it is difficult to extract sufficient numbers of control point correspondences, and its ability to consider any desired set of control point extraction algorithms. The resulting registration schemes had more control point correspondences on average than the traditional register-all-to-one-target-channel approach in all of our experiments. Registration schemes showed consistently lower back-projection error than traditional direct-to-target-channel based registration approach.

The overall success of data-driven image registration process depends on the performance of the image registration control point extraction algorithms to find out the maximum number of control point correspondence with a minimum back-projection error. In this research, we have considered widely used standard control point extraction algorithms SIFT, SURF, BRISK, ORB. Our experiment on canola crop breeding multispectral image dataset showed that SIFT is the superior control point extraction algorithm based on average control-points, back-projection error.

Data-driven multispectral image registration framework is applicable to any multispectral dataset. It is registering multispectral images with any number of image channels of any multispectral cameras. Our data-driven framework is generating a registration scheme with standard control point extraction algorithms and registering images in a data-driven approach where large numbers of control-point matches is our priority. With the growing need for remote sensing and lack of proper multispectral image registration technique, data-driven registration framework will be proven as an essential tool in the field of image registration and multispectral imaging.

7.1 Future work

We have tested our data-driven multispectral image registration framework in five different multispectral image dataset collected using three different cameras. Three of these datasets have five image channels, and two of them have four image channels. We want to test data-driven registration framework in a dataset where there are more than five image channels available.

The data-driven image registration process goes through all the images to collect average control point correspondence and then generatesithe registration scheme. After producing the registration scheme, we are regisng multispeal images. As we are going through all the images of the entire dataset or, a subset of images of the dataset, takesit is taking morstandardthe general image r approaches. We want to optimize the overall process of registration of data-driven approach by running registration scheme generation process an image registration process in a multi-threaded process using parallel processing.

The data-driven multispectral image registration framework goes through all the images in a dataset to get desired registration scheme. In the future, our framework should be able to select a fraction of dataset by itself to build the registration scheme incrementally instead of checking all the images in the dataset. Multispectral image registration process will be much faster in this way.

In this proposed research we used perspective image transformation using a homography matrix to solve the misalignment issue. Instead of using a homography matrix to transform images, we could apply a convolutional neural network technique to align multichannel images according to the root channel obtained by the data-driven process.

The data-driven multispectral image registration framework is generating dataset specific registration scheme using average control point correspondences. We are evaluating the performance of the registration using back-projection error. In the future, we would like to introduce more methods of evaluation for the performance of registration scheme.

The data-driven multispectral image registration framework is the only available multi-spectral image registration solution now. We may launch this framework as an extensible software with a plug-in system for more control point algorithms and as a handy tool for any multispectral image registration dataset.

References

[1] (2011-2014) Back Projection Calculation, OpenCV 2.4.13.5 documentation. [On-line]. https://docs.opencv.org/2.4/doc/tutorials/imgproc/histograms/back_projection/back_projection.html.

[2] (2011-2014) Warp Perspective Calculation, Opencv 2.4.13.5 documentation. [On-line]. https://docs.opencv.org/3.0-beta/doc.html.

[3] (2015) TokyoTech 5-band Multispectral RAW Dataset. [On-line]. http://www.ok.sc.e.titech.ac.jp/res/MSI/MSIdata31.html.

[4] (2017) Assessing Crop Multispectral eBee Ag and SQ Dataset. [On-line]. https://www.sensefly.com/drones/example-datasets.html.

[5] (2018) Mica RedEdge Camera specification. [On-line]. https://www.micasense.com/rededge/.

[6] Alexandre Alahi, Raphael Ortiz, and Pierre Vandergheynst. FREAK: Fast retina keypoint. *IEEE Conference on Computer Vision and Pattern Recognition (2012)*, pages 1–6, 2012.

[7] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110:346–359, 2008.

[8] Somaraju Boda. Feature-based image registration. Master's thesis, National Institute of Technology, Rourkela, 2009.

[9] Guo-Rong Cai, Pierre-Marc Jodoin, Shao-Zi Li, Yun-Dong Wu, Song-Zhi Su, and Zhen-Kun Huang. Perspective-SIFT: An efficient tool for low-altitude remote sensing image registration. *Signal Processing, 2013 Elsevier B.V.*, 93:3088–3110, 2013.

[10] Michael Calonder, Vincent Lepetit, Christoph Strecha, and Pascal Fua. BRIEF: Binary robust independent elementary features. *11th European Conference on Computer Vision (ECCV), Heraklion, Crete. LNCS Springer, September 2010*, pages 778–792, 2010.

[11] Elan Dubrofsky. Homography estimation. Master's thesis, The University of British Columbia, 2009.

[12] D. Firmenich, M. Brown, and S. Süsstrunk. Multispectral interest points for RGB-NIR image registration. *18th IEEE International Conference on Image Processing*, pages 181–184, 2011.

[13] Robert W. Floyd. Algorithm 97: Shortest path. *Communications of the ACM*, 5(6):345, 1962.

[14] Gang Hong and Yun Zhang. Wavelet-based image registration technique for high-resolution remote sensing images. *Computers and Biomedical ResearchGeosciences, 2008 Elsevier Ltd*, 34:1708–1720, 2008.

[15] Sahin Işik and Kemal Özkan. A comparative evaluation of well-known feature detectors and descriptors. *International Journal of Applied Mathematics, Electronics and Computers, ISSN: 2147-8228*, 3(1):1–6, August 2014.

[16] Luo Juan and Oubong Gwun. A comparison of SIFT, PCA-SIFT and SURF. *International Journal of Image Processing*, 3(4):143–152, 2011.

[17] Jude H. Kastens, Terry L. Kastens, Dietrich L.A. Kastens, Kevin P. Price, Edward A. Martinko, and Re-Yang Lee. Image masking for crop yield forecasting using AVHRR NDVI time series imagery. *Remote Sensing of the Environment*, 9(3):341–356, 2005.

[18] Yan Ke and Rahul Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, 2:II–506–II–513, 2004.

[19] Jeffrey P. Kern and Marios S. Pattichis. Robust multispectral image registration using mutual-information models. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5):1494–1505, 2007.

[20] J. B. Kruskal. On the shortest spanning subtree of a graph and the travelling salesman problem. *Proceedings of the American Mathematical Society*, 7:48–50, 1956.

[21] Paula Beatriz Cerqueira Leite, Raul Queiroz Feitosa, Antônio Roberto Formaggio, Gilson Alexandre Ostwald Pedro da Costa, Kian Pakzad, and Ieda Del'Arco Sanches. Hidden markov models for crop recognition in remote sensing image sequences. *Pattern Recognition Letters, 2010 Elsevier B.V.*, 32:19–26, 2011.

[22] Stefan Leuteneggar, Margaria Chli, and Roland Y. Siegwart. BRISK: Binary robust invariant scalable keypoints. *International Conference on Computer VIsion 2011*, pages 2584–2555, 2011.

[23] Douglas Lim. Achieving accurate image registration as the basis for super-resolution. Master's thesis, The University of Western Australia,, 2003.

[24] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 2004*, 60(2):91–110, 2004.

[25] Vladimir V. Riabov. Exploring algorithms for effective applications of the graph theory. In *26th International Conference on Technology in Collegiate Mathematics*, 2014.

[26] Ibrahim A. El Rube, Maha A. Sharks, and Ashor R. Salem. Image registration based on multi-scale sift for remote sensing images. *IEEE Conference 978-1-4244-4474-8/09*, pages 1–5, 2009.

[27] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: an efficient alternative to SIFT or SURF. *IEEE International Conference on Computer Vision (ICCV)*, pages 2564–2571, 2011.

[28] P. Schwind and P. d'Angelo. Evaluating the applicability of BRISK for the geometric registration of remote sensing images. *Remote Sensing Letters*, 6(9):677–686, July 2015.

[29] Mustafa Teke and Alptekin Temizel. Multi-spectral satellite image registration using scale-restricted SURF. *2010 International Conference on Pattern Recognition*, pages 2310–2313, 2010.

[30] Xuezhi Wang, Weiping Yang, Ashley Wheaton, Nicola Cooley, and Bill Moran. Efficient registration of optical and IR images for automatic plant water stress assessment. *Computers and Electronics in Agriculture*, 74:230–237, 2010.

[31] Z. Yi. C. Zhiguo and X. Yang. Multi-spectral remote image registration based on SIFT. *Electronics Letters online no: 20082477*, 44(2):107–108, 2008.

[32] Yingxuan Zhu, Samuel Cheng, Vladimir Stankovic , and Lina Stankovic. Image registration using BP-SIFT. *J. Vis. Commun. Image R*, 24:448–457, 2013.

[33] Barbara Zitová and Jan Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21:977–1000, 2003.